COSMICAL
ELECTRO-
DYNAMICS

ALFVÉN

OXFORD

THE

INTERNATIONAL SERIES OF MONOGRAPHS ON PHYSICS

GENERAL EDITORS

†R. H. FOWLER, P. KAPITZA
N. F. MOTT, E. C. BULLARD

THE INTERNATIONAL SERIES OF MONOGRAPHS ON PHYSICS

GENERAL EDITORS

Already Published

THE THEORY OF ELECTRIC AND MAGNETIC SUSCEPTIBILITIES. By J. H. VAN VLECK. 1932.

THE THEORY OF ATOMIC COLLISIONS. By N. F. MOTT and H. S. W. MASSEY. *Second edition.* 1949.

RELATIVITY, THERMODYNAMICS, AND COSMOLOGY. By R. C. TOLMAN. 1934.

CHEMICAL KINETICS AND CHAIN REACTIONS. By N. SEMENOFF. 1935.

RELATIVITY, GRAVITATION, AND WORLD-STRUCTURE. By E. A. MILNE. 1935.

KINEMATIC RELATIVITY. A sequel to *Relativity, Gravitation, and World-Structure.* By E. A. MILNE. 1948.

THE QUANTUM THEORY OF RADIATION. By W. HEITLER. *Second edition.* 1945.

THEORETICAL ASTROPHYSICS: ATOMIC THEORY AND THE ANALYSIS OF STELLAR ATMOSPHERES AND ENVELOPES. By S. ROSSELAND. 1936.

ECLIPSES OF THE SUN AND MOON. By SIR FRANK DYSON and R. V. d. R. WOOLLEY. 1937.

THE PRINCIPLES OF STATISTICAL MECHANICS. By R. C. TOLMAN. 1938.

ELECTRONIC PROCESSES IN IONIC CRYSTALS. By N. F. MOTT and R. W. GURNEY. *Second edition.* 1948.

GEOMAGNETISM. By S. CHAPMAN and J. BARTELS. 1940. 2 vols.

THE SEPARATION OF GASES. By M. RUHEMANN. *Second edition.* 1949.

KINETIC THEORY OF LIQUIDS. By J. FRENKEL. 1946.

THE PRINCIPLES OF QUANTUM MECHANICS. By P. A. M. DIRAC. *Third edition.* 1947.

THEORY OF ATOMIC NUCLEUS AND NUCLEAR ENERGY SOURCES. By G. GAMOW and C. L. CRITCHFIELD. 1949. *Being the third edition of* STRUCTURE OF ATOMIC NUCLEUS AND NUCLEAR TRANSFORMATIONS.

THE PULSATION THEORY OF VARIABLE STARS. By S. ROSSELAND. 1948.

COSMIC RAYS. By L. JÁNOSSY. *New edition in preparation.*

THEORY OF PROBABILITY. By HAROLD JEFFREYS. *Second edition.* 1948.

RADIATIVE TRANSFER. By S. CHANDRASEKHAR. 1950.

COSMICAL ELECTRODYNAMICS

BY

H. ALFVÉN
PROFESSOR OF ELECTRONICS,
ROYAL INSTITUTE OF TECHNOLOGY,
STOCKHOLM

OXFORD
AT THE CLARENDON PRESS
1950

Oxford University Press, Amen House, London E.C.4
GLASGOW NEW YORK TORONTO MELBOURNE WELLINGTON
BOMBAY CALCUTTA MADRAS CAPE TOWN
Geoffrey Cumberlege, Publisher to the University

PRINTED IN GREAT BRITAIN

PREFACE

Recent discoveries have revealed that electromagnetic phenomena are of greater importance in cosmic physics than used to be supposed. The time now seems to be ripe for an attempt to trace systematically the electromagnetic phenomena in the cosmos, and this is the reason for writing the present volume.

Cosmic physics is still in the stage where the most important task is to find out what are the dominating physical factors. Too many theories have been worked out with much mathematical skill on basic assumptions which were not physically tenable. Hence in this book the stress is always laid more on the physical than on the mathematical side. It is clearly understood that definite tests of any theory can be made only by means of rigorous mathematics, but the scope of this book is more to put the problems than to solve them.

The first four chapters are of fundamental character, the last three contain the applications. The reader is supposed to be familiar with the empirical results in this field. No attempt has been made to give an historical account of the development of the theories.

During a prolonged correspondence and many discussions Mr. Nicolai Herlofson has offered most valuable criticism from which I have profited. My thanks are also due to Mr. Stig Lundquist who has very kindly helped me with the preparation of the manuscript.

STOCKHOLM
THE ROYAL INSTITUTE OF TECHNOLOGY
July 1948

CONTENTS

I

GENERAL SURVEY

1.1. PHYSICS is mainly based on experience gained in the laboratory. When we try to apply to cosmic phenomena the laws in which this experience is condensed, we make an enormous extrapolation, the legitimacy of which can be checked only by comparing the theoretical results with observations. Classical mechanics was once extrapolated into the realm of astronomy so successfully that only the most refined observations of the last decades have revealed phenomena for which it does not hold. The application of atomic theory, especially spectroscopy, to cosmic phenomena has proved equally successful. In fact, classical mechanics and spectroscopy have been two invaluable tools in exploring the universe around us.

It seems very probable that electromagnetic phenomena will prove to be of great importance in cosmic physics. Electromagnetic phenomena are described by classical electrodynamics, which, however, for a deeper understanding must be combined with atomic physics. This combination is especially important for the phenomena occurring at the passage of current through gaseous conductors which are treated by the complicated theory of 'discharges' in gases. No definite reasons are known why it should not be possible to extrapolate the laboratory results in this field to cosmic physics. Certainly, from time to time, various phenomena have been thought to indicate that ordinary electrodynamic laws do not hold for cosmic problems. For example, the difficulty of accounting for the general magnetic fields of celestial bodies has led different authors, most recently Blackett (1947), to assume that the production of a magnetic field by the rotation of a massive body is governed by a new law of nature. If this is true, Maxwell's equations must be supplemented by a term which is of paramount importance in cosmic physics. The arguments in favour of a revision are still very weak. Thus it seems reasonable to maintain the generally accepted view that all common physical laws hold up to lengths of the order of the 'radius of the universe' and times of the order of the 'age of the universe', limits given by the theory of general relativity.

The discovery of sunspot magnetic fields (Hale, 1908) and later of the sun's general field (Hale, Seares, von Maanen, and Ellerman, 1918) has been of decisive importance to cosmic electrodynamics. More recently Babcock (1947) has shown that even stars possess strong magnetic

fields. It may be said that if the sun and stars had no magnetic fields, electromagnetic phenomena would be of little importance to cosmic physics.

Celestial magnetic fields affect the motion of charged particles in space. Under certain conditions electromagnetic forces are much stronger than gravitation. In order to illustrate this, let us suppose that a particle moves at the earth's solar distance $R_{♁}$ with the earth's orbital velocity $v_{♁}$. If the particle is a neutral atom, it is acted upon only by the solar gravitation (the effect of the solar magnetic field upon an eventual atomic magnetic moment being negligible). If $M_{\odot}$ is the solar and m_A the atomic mass, and k is the constant of gravitation, this force is

$$f_A = kM_{\odot}\, m_A/R_{♁}^2.$$

If the atom becomes singly ionized, the ion as well as the electron (charge $= \pm e$) is subject to the force

$$\mathbf{f}_m = (e/c)[\mathbf{v}_{♁}\,\mathbf{H}_{♁}]$$

from the solar magnetic field $\mathbf{H}_{♁}$. Under the assumption that this field is due to a dipole with the moment $a = 0{\cdot}42\,.\,10^{34}$ gauss cm.^{3}, we find $H_{♁} = 1{\cdot}2\,.\,10^{-6}$ gauss. If m_A is the mass of a hydrogen atom it is easily found that

$$f_m/f_A \sim 10^5.$$

This illustrates the enormous importance of the solar magnetic field even at the earth's distance from the sun.

On the other hand, as f_m has opposite signs for electrons and for ions, in many cases the forces on electrons and ions may cancel each other. If we substitute for the particle an ionized cloud, containing the same number of electrons and ions, the resulting magnetic force on the cloud becomes zero to a first approximation. Second-order effects, e.g. due to the inhomogeneity of the magnetic field, may still be important. Further, the motion in the magnetic field produces a separation of the ions and electrons, but the resulting polarization causes an electric field which limits the separation. Under certain conditions the electric field may produce currents in adjacent conductors so that very complicated phenomena occur.

In the sun itself the magnetic field is of importance in several respects. In the outer layers, the chromosphere and the corona, the radius of curvature of the path of a charged particle with thermal velocity is smaller than the mean free path. Hence the magnetic field introduces an anisotropy, so that, for example, the electric conductivity is higher

in the direction of the magnetic field than perpendicular to it (Cowling, 1932). In the photosphere, and in the sun's interior, the mean free path is small in comparison to the radius of curvature, which means that solar matter can be treated as an isotropic conductor. But even in this case an anisotropy is introduced by the fact that currents perpendicular to the magnetic field produce forces which accelerate the medium. A consequence of this is that magneto-hydrodynamic waves (see Chap. IV) move in the direction of the magnetic field.

The examples above demonstrate on the one hand the importance of electromagnetic forces in cosmic physics, and on the other the complexity of the electromagnetic phenomena. In our attempt to trace electromagnetic effects we shall start with a discussion of the magnetic and electric fields in cosmic physics. In Chapter II we shall treat the motion of a single particle in such fields. If several charged particles are present, forming an ionized gas, phenomena related to those studied in electric discharges are likely to occur. A survey of these phenomena is given in Chapter III. At densities so great that the ionized gas can be considered as an ordinary electrical conductor, the most important phenomenon in connexion with electromagnetic forces is probably that of magneto-hydrodynamic waves. These are treated in Chapter IV.

The results are applied to solar physics in Chapter V and to the theory of magnetic storms and aurora in Chapter VI. A discussion of the astrophysical aspect of cosmic radiation is given in Chapter VII.

It was originally intended to discuss an electromagnetic theory of the origin of the solar system (Alfvén, 1942, 1943, 1946) in an eighth chapter. This has been excluded, however, because it would require rather too much space.

It is a matter of judgement whether the physics of the ionosphere should be reckoned as cosmic electrodynamics or not. Certainly it has close connexions with, for example, the theory of the solar corona. On the other hand, it is still more closely related to the extensive field of the physics of the upper atmosphere. As even a superficial treatment of these problems would require too much space, the physics of the ionosphere has been excluded altogether.

1.2. Magnetic fields in cosmic physics

Every electric current, and what is equivalent to that, every magnet, gives a magnetic field which at great distances approximates to a dipole field. Hence in the absence of currents in the surroundings the fields of the earth and the sun are dipole fields at great distances. For the

earth, and probably also for the sun, this approximation is rather close even at the surface, and hence everywhere above the surface.

A dipole with moment a situated at the origin and parallel to the z-axis

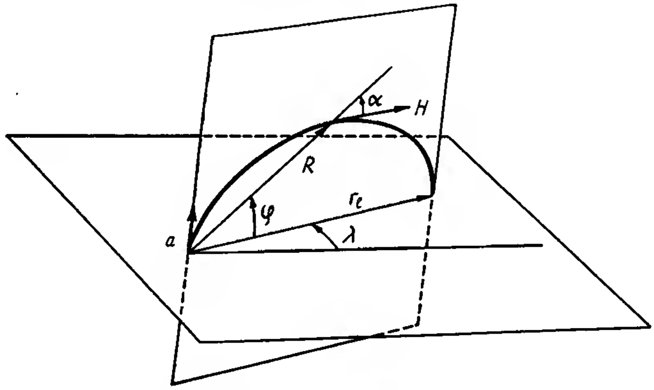

FIG. 1.1. Magnetic line of force from a dipole a.

gives a field, the components of which in a spherical coordinate system (R, φ, λ) are

$$H_R = H_p \sin\varphi, \tag{1}$$

$$H_\varphi = -\tfrac{1}{2}H_p \cos\varphi, \tag{2}$$

$$H_\lambda = 0,$$

$$H_p = 2a/R^3, \tag{3}$$

$$H = \sqrt{(H_R^2+H_\varphi^2+H_\lambda^2)} = a\phi/R^3 = \tfrac{1}{2}H_p\,\phi, \tag{4}$$

where

$$\phi = \sqrt{(1+3\sin^2\varphi)}. \tag{5}$$

H_R represents the 'vertical' and H_φ the 'horizontal' component of the field.† A magnetic line of force has the equation

$$R = r_e \cos^2\varphi, \tag{6}$$

$$\lambda = \text{const.},$$

where r_e is the distance from the origin of the point where it intersects the equatorial plane ($\varphi = 0$). The angle α between the line of force and the radius vector is given by

$$\tan\alpha = \tfrac{1}{2}\cot\varphi, \tag{7}$$

or

$$\sin\alpha = \frac{\cos\varphi}{\phi}, \tag{8}$$

$$\cos\alpha = \frac{2\sin\varphi}{\phi}. \tag{9}$$

The 'inclination' of the field is $\tfrac{1}{2}\pi-\alpha$.

† In terrestrial magnetism the 'vertical component' is counted positive if directed downwards.

The total strength of the field along a given line of force can also be written

$$H = \frac{a}{R^3}\phi = \frac{a}{r_e^3}(\cos\varphi)^{-6}\phi = \frac{a}{r_e^3}\eta, \tag{10}$$

where

$$\eta = \frac{\sqrt{(1+3\sin^2\varphi)}}{\cos^6\varphi}. \tag{11}$$

In a Cartesian system (x, y, z) we have

$$H_x = 3xz\frac{a}{R^5}, \tag{12}$$

$$H_y = 3yz\frac{a}{R^5}, \tag{13}$$

$$H_z = (3z^2-R^2)\frac{a}{R^5} \tag{14}$$

with $R^2 = x^2+y^2+z^2$.

If the *terrestrial field* is treated as the field from a dipole situated at the earth's centre, this dipole has the moment (see Chapman and Bartels, 1940, p. 645)

$$8{\cdot}1\,.\,10^{25} \text{ gauss cm.}^3 \tag{15}$$

Its axis intersects the earth's surface in two antipodal points situated at latitude 78·5° S., longitude 111° E., and at latitude 78·5° N., longitude 69° W. The dipole moment (15) corresponds according to (3) to $H_p = 0{\cdot}63$ gauss.

A better approximation is obtainable if the condition that the dipole should be situated at the centre is dropped. The best agreement with the real field is obtained if the dipole is shifted 342 km. from the centre towards the point 6·5° N., 161·8° E. The axis of the eccentric dipole intersects the earth's surface at two points, 76·3° S., 121·2° E., and 80·1° N., 277·3° E.

The terrestrial field is subject to a slow (secular) variation. At present the magnetic moment seems to decrease by about 0·1 per cent. per year.

The *solar magnetic field* has been determined by measuring the Zeemann effect. The displacement of the sunspot zone (see § 5.31) and some other effects supply additional, although less direct, arguments for the existence of a general magnetic field. The properties of the field are discussed in § 5.22. The polar strength is likely to be about 25 gauss, corresponding to a dipole moment of

$$4{\cdot}2\,.\,10^{33} \text{ gauss cm.}^3 \tag{16}$$

Because of the difficulty of exact measurements, this value may be in error by a factor 2, perhaps even more.

Sunspots are always associated with strong magnetic fields, as big as 4,000 gauss.

Stellar magnetic fields have been discovered by Babcock (1947) through Zeemann effect measurements. For 78 Virginis he finds a polar strength of 1,500 gauss corresponding to a moment of 4.10^{36} gauss cm.^{3}, and for the star BD 18°3789 (HD 1252 48) the field is no less than 5,500 gauss. The field of the latter object seems to be variable.

There are some arguments for the existence of a general *galactic magnetic field*. This problem is treated in § 7.5.

1.3. Induced electric field

In the presence of a magnetic field an electric field is defined only in relation to a certain coordinate system. If in a system 'at rest' the electric and magnetic fields are **E** and **H**, we can calculate by means of relativistic transformation formulae the fields **E**$'$, **H**$'$, in a system which moves in relation to the first with the velocity **v**. The components parallel to **v** remain unchanged, but the components perpendicular to **v** are transformed in the following way:

$$\mathbf{E}' = \frac{\mathbf{E}+c^{-1}[\mathbf{vB}]}{\sqrt{(1-c^{-2}v^2)}}, \tag{1}$$

$$\mathbf{H}' = \frac{\mathbf{H}-c^{-1}[\mathbf{vD}]}{\sqrt{(1-c^{-2}v^2)}} \tag{2}$$

($D = \epsilon E$, $B = \mu H$; reduced in a vacuum to $D = E$, $B = H$).

The astronomical velocities are much smaller than the velocity of light (c). Because of the good conductivity, electrostatic fields will usually be of little importance. Hence the electric fields are usually secondary to the magnetic fields, which, according to (1), means that the electric fields are much weaker than the magnetic fields. Consequently in cosmic physics we can usually to a good approximation write

$$\mathbf{E}' = \mathbf{E}+(1/c)[\mathbf{vH}], \tag{3}$$

$$\mathbf{H}' = \mathbf{H} \tag{4}$$

(where also the components parallel to v are included in the vectors).

Thus the magnetic fields are independent of the coordinate system, but *to speak of an electric field without defining exactly the coordinate system to which it refers is meaningless.*

These simple and fundamental principles seem to have attracted very little interest from astrophysicists and geophysicists. They are not very much to blame because the subject is omitted in most treatises on

electromagnetism. Formulae (1) and (2) are found in books on the theory of relativity, e.g. Riemann–Weber (1927) and McCrea (1935).

The importance of the relativity of electric fields in cosmic physics is enormous. One of the consequences is that all celestial bodies with magnetic fields are on account of their rotation electrically polarized when seen from a system at rest. This phenomenon is well known from laboratory experiments and is usually called 'homopolar' or 'unipolar' induction. It was first studied by Faraday, and attracted much interest during the last century because it was thought that by investigating this subject it should be possible to ascertain whether the magnetic lines of force from a rotating magnet rotate with the magnet or not. At present the phenomenon seems to be half-forgotten, and most text-book authors do not mention it. A noteworthy exception is Cullwick (1939), who devotes a special appendix to it. It is also discussed by Becker (1933). Further, it should be mentioned that one of the best methods for absolute determination of the ohm employs a unipolar inductor (see Curtis, 1937). The device has also been developed electrotechnically as a direct current generator producing currents of thousands of amperes (see, for example, Arnold–la Cour, 1919).

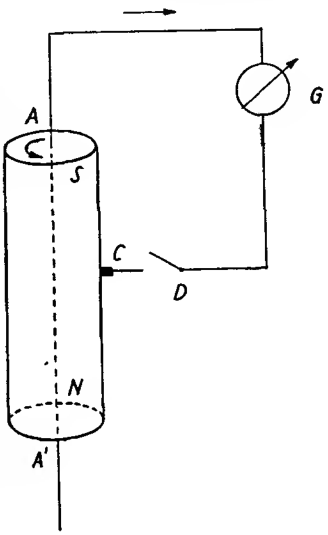

FIG. 1.2. Unipolar inductor. When a bar-magnet N–S is rotated around its axis AA' a current is obtained in a *fixed* circuit connecting the axis with a sliding contact at C.

A simple unipolar inductor is obtained by rotating a cylindrical bar-magnet N–S around its axis AA' (see Fig. 1.2). A fixed wire $AGDC$ connects the axis with a sliding contact C at the middle of the bar. If the switch D is closed the galvanometer G indicates a current as soon as the magnet rotates. The phenomenon can be treated *either* in a fixed system *or* in a rotating system. In the first case the magnetic lines of force are considered to be at rest *outside as well as inside* the magnet. The motion of the magnet produces a polarization inside it so that positive charge is accumulated near the axis and negative charge near the sliding contact. If the circuit is interrupted, this accumulation proceeds until the field from the charges neutralizes the polarization field, so that the resulting field E' becomes zero. This is necessary because

the magnet is a conductor and when the current is zero the electric field seen from a system moving with a conductor must be zero. Then we have from (3)

$$\mathbf{E} = -(1/c)[\mathbf{vB}]. \tag{5}$$

The voltage difference between the sliding contact and the axis is

$$V = -(1/c)\int_A^C [\mathbf{vB}]\,d\mathbf{s}, \tag{6}$$

where $d\mathbf{s}$ is a line element.

If the switch is closed this voltage produces a current in the circuit.

This discussion is founded on the assumption that the magnetic field is 'at rest'. The problem can also be treated under the assumption that the lines of force take part in the rotation, and the result is the same. In this case no polarization is produced inside the magnet, but outside the magnet the wire $AGDC$ constantly cuts magnetic lines of force. Hence an e.m.f. is induced and it is easily shown that this has the value (6).

When we treat a problem in the rotating system we must observe that according to the general theory of relativity the electrodynamic equations for a rotating system do not have the usual form. In the presence of a magnetic field B the electric field deriving from a space charge ρ can be found from

$$4\pi\rho = \operatorname{div} E - (2/c)(\boldsymbol{\omega}\mathbf{B}),$$

where ω is the angular velocity.† Within a conductor we have $E = 0$, and the space charge is given by

$$\rho = -(\boldsymbol{\omega}\mathbf{B})/2\pi c.$$

The same result can also be obtained in the fixed system by taking the divergence of (5).

If the magnet is surrounded by an ideal insulator, we have outside the magnet $\rho = 0$. Hence, even in the rotating system, an electric field is produced. If, on the other hand, the surrounding medium has a conductivity which differs from zero, we have also $E = 0$ outside the magnet. In this case the magnetic lines of force may be considered as rotating with the magnet. As we have assumed $v \ll c$, the result does not, of course, hold for large distances from the axis.

It is not essential that the rotating body should be a permanent magnet. Any conductor will do if only a magnetic field is established in some way. In Fig. 1.3 a coil produces the magnetic field in which a copper disc rotates around the axis AA'. The e.m.f. is given by (6).

† I am indebted to Professor O. Klein for pointing this out to me.

It should be observed that if an instrument G' is placed on the disc, so that it takes part in the rotation and is connected between the axis and periphery, the voltage zero is read on this instrument.

After having discussed various types of earthly unipolar inductors we may be allowed to extrapolate the results to cosmic phenomena. It is obvious that the earth and the sun must be polarized in the same way as the bar-magnet or the copper disc. Let us consider the fields of these bodies as dipole fields with the magnetic axis coinciding with the rotational axis and neglect the non-uniform rotation of the sun. Because of the good electric conductivity the electrostatic potential must be the same at the poles as at the equator, *when measured in a system which takes part in the rotation*. Transforming to a system at rest (not partaking in the rotation) the bodies are electrically polarized according to (3). The electric field lies in the meridian plane and its horizontal component E_1 amounts to

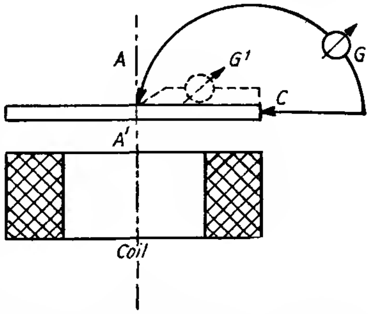

FIG. 1.3. Unipolar inductor, consisting of a rotating copper disc which is polarized by the field from a coil.

$$vH_R/c,$$

where H_R is the vertical component of the magnetic field. Putting

$$v = v_e \cos\varphi,$$
$$H_R = H_p \sin\varphi$$

(φ = latitude, v_e = equatorial velocity, H_p = polar field strength), we obtain

$$E_1 = \frac{v_e H_p}{2c} \sin 2\varphi. \tag{7}$$

For the earth we have $v_e = 0{\cdot}5 \,.\, 10^5$ cm. sec.$^{-1}$, $H_p = 0{\cdot}6$ gauss. Hence we obtain for $\phi = 45°$, the field $E_1 = 0{\cdot}5 \,.\, 10^{-6}$ e.s.u. $= 150\ \mu$ volt cm.$^{-1}$ Integrating (7) from the equator to the pole, we find that the voltage difference between equator and pole is given by $V = \int E\, ds = 10^5$ volts if *measured from a system at rest*.

In the same way we find that *seen from a system at rest* there is a voltage difference between the equator and the poles of the sun of $1{\cdot}7 \,.\, 10^9$ volts. As in the case of the earth, the equator is negative in relation to both poles.

The surface charge of the rotating body produces an electric quadrupole field outside the body. In the case of a rotating sphere this field has been calculated by Davis (1947). If the body is surrounded by a conducting medium, the electric field is modified so that it becomes

$$E = vH/c,$$

with $v = r\omega$, which means that the surrounding medium tends to share the rotational state of the body.

The consequences of the unipolar action of celestial bodies will be discussed further in §§ 5.61 and 5.82.

Another example of unipolar induction is found in the solar atmosphere, where motions in the general magnetic field or sunspot fields may produce very large electromotive forces (see § 5.61).

When an ionized cloud moves in a magnetic field it becomes polarized according to (3). For example, the ionized clouds, which according to current ideas of the cause of magnetic storms are emitted from the sun (see § 6.1), are electrically neutral when seen from a coordinate system which moves with the cloud. When seen from the earth, which in this connexion may be considered as approximately at rest, they are electrically polarized (compare Becker, p. 336). As we shall see in Chapter VI, this field is probably of decisive importance in the theory of magnetic storms and aurorae.

For the production of electric fields according to (3) we use the term *polarization* or (with Cullwick) *motional induction*. Unipolar induction is one special case of this and the polarization of an ionized cloud another.

Electric fields may also be produced by a change in magnetic field. This type of induction is called by Cullwick *transformer induction*. The field may be calculated from Maxwell's equation

$$\operatorname{curl} \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}$$

or from

$$\mathbf{E} = -\frac{d\mathbf{A}}{dt},$$

where $\mathbf{B} = \mu\mathbf{H}$ is the magnetic flux density and $\mathbf{A}$ the vector potential.

In cosmic physics large electromotive forces are produced either by motional induction or transformer induction. In special cases electrostatic fields must also be taken into consideration. Small voltage differences, due to diffusion, thermo-electric or electro-chemical effects, may in some cases be of importance.

1.4. Approximate equality of positive and negative space charge

Consider a sphere of radius R containing N_1 positive charges e and N_2 negative charges $-e$ per unit volume. The electrostatic potential at its surface is

$$V = \frac{4\pi}{3}(N_1-N_2)eR^2.$$

When we deal with a problem in, for example, the solar corona, we can be sure that there cannot be a potential of say 3.10^{10} volts ($= 10^8$ e.s.u.). A sphere with radius R equal to 10^9 cm. is so small a part of the corona that all densities are likely to be approximately uniform. Inserting the value of R and putting $e = 4{\cdot}8.10^{-10}$ e.s.u., the condition $V < 10^8$ gives

$$N_1-N_2 < 0{\cdot}05 \text{ particles.cm.}^{-3}$$

As $N_2 \sim 10^8$ cm.$^{-3}$, we find

$$\frac{N_1-N_2}{N_2} < 0{\cdot}5.10^{-9}.$$

Hence even if there are only 10^9+1 electrons for 10^9 protons an impossibly high voltage is produced.

Similar results are obtained for almost all other cosmic problems.

Hence we may conclude that *in cosmic physics the positive space charge in a volume is always approximately equal to the negative space charge.*

From the study of electric discharges in gases it is also known that the number of positive and of negative particles must be approximately the same, as soon as the charged particle density is large.

REFERENCES

Alfvén, H., 1942, 1943, 1946: 'On the Cosmogony of the Solar System, I', *Stockholms Observatoriums Annaler*, I, **14**, No. 2, 1942; II, ibid., **14**, No. 5, 1943; III, ibid., **14**, No. 9, 1946.

Arnold–la Cour, 1919: *Die Gleichstrommaschine*, **1**, p. 1, Berlin.

Babcock, H. W., 1947: 'Remarks on Stellar Magnetism', *Publ. Astr. Soc. Pacific*, **59**, 112, No. 348; 'The Reversing Magnetic Field of BD–18°3789', ibid., **59**, 260, No. 350.

Becker, R., 1933: *Theorie der Elektrizität* II, *Elektronentheorie*, p. 334, Loipzig.

Blackett, P. M. S., 1947: 'The Magnetic Field of Massive rotating Bodies,' *Nature*, **159**, 658.

Chapman, S., and Bartels, J., 1940: *Geomagnetism*, pp. 639 and 668.

Cowling, T. G., 1932: 'The Electrical Conductivity of an Ionized Gas in the Presence of a Magnetic Field', *Mon. Not. R. Astr. Soc.*, **93**, 90. See also Chapman, S. and Cowling, T. G., 1939: *The Mathematical Theory of Non-Uniform Gases*, p. 322, Cambridge.

CULLWICK, E. G., 1939: *The Fundamentals of Electromagnetism*, p. 321, Cambridge.

CURTIS, H. L., 1937: *Electrical Measurements*, p. 196, New York.

DAVIS, L., 1947: 'Stellar Electromagnetic Fields', *Phys. Rev.*, **72**, 632.

HALE, G. E., 1908: 'On the Probable Existence of a Magnetic Field in Sun-spots', Mt. Wilson Contr. No. 30, *Astrophys. J.*, **28**, 315.

——, SEARES, F. H., VON MAANEN, A., and ELLERMAN, F., 1918: 'The General Magnetic Field of the Sun', Mt. Wilson Contr. No. 148, ibid. **47**, 206.

McCREA, W. H., 1935: *Relativity Physics*, p. 51, Methuen's Monographs.

RIEMANN–WEBER, 1927: Riemann–Weber's *Differentialgleichungen der Physik*, **2**, p. 410 (7th edition, Braunschweig, 1927).

II

ON THE MOTION OF CHARGED PARTICLES IN MAGNETIC FIELDS

2.1. THE first to appreciate fully the paramount importance to cosmic physics of the problems of the motion of charged particles in magnetic fields were Birkeland and Störmer. Inspired by Birkeland's terrella experiment (see § 6.1), Störmer has devoted a long series of papers especially to motion in a magnetic dipole field. He has also discussed the effect of an additional electric field. The problems are treated by exact mathematical methods, but the final results about the paths of the particles can be obtained only by very laborious numerical integrations. It has not been possible to carry through the calculations except for the case when the particle does not make too many loops in the magnetic field. This means that Störmer's method is generally applicable only for particles above a certain momentum. In the terrestrial field the limit lies in the range of cosmic rays. Of the trajectories corresponding to momenta below this only some special types (trajectories through the dipole) can be found with a reasonable amount of labour. In the general case particles below the cosmic-ray range make hundreds or thousands of loops in the terrestrial magnetic field, and unless modern calculating machines are used the integration is practically impossible.

In the case when the path makes many loops the size of one loop is in general small compared with the extension of the magnetic field. Hence during a single turn the particle moves in an approximately homogeneous field. In order to calculate the motion it is advantageous to start with the motion in a homogeneous field and introduce the inhomogeneity as a perturbation (Alfvén, 1940). This *perturbation method* which is developed in § 2.2 and § 2.3 is especially suited for low-energy particles. In the terrestrial field it is applicable to almost all problems where the momentum is below the cosmic-ray range. Thus the two methods seem to complement each other.

The paths may also be found by direct experiment. Brüche (1931) produced very thin electron beams and studied their paths in the dipole field from a homogeneously magnetized iron sphere. The method has been further developed by Malmfors (1945), who has determined quantitatively a number of orbits which are of interest in cosmic-ray problems. Malmfors's diagrams give a very valuable survey of the motion of cosmic rays in the terrestrial field.

After a survey of the motion in a homogeneous field we shall develop the perturbation method of calculating the path in a field of arbitrary form (§§ 2.2 and 2.3). The motion in a dipole field is treated by Störmer's method (§ 2.4), as well as by the perturbation method (§ 2.5), and the results are compared.

2.2. Homogeneous magnetic field

In this paragraph we shall first give some simple formulae referring to the motion of charged particles in a homogeneous magnetic field and then discuss the perturbation method, which will be further developed in § 2.3. This method makes use of the fact that when a particle which spirals in a magnetic field is acted upon by a force, it moves ('drifts') perpendicular to the force. This is expressed by equation (24). To the applied force f an eventual inertia force f^i should be added. When the magnetic field is inhomogeneous a fictitious force f^m should be introduced leading to the more general equation (32) of § 2.3. A simple physical interpretation of f^m is possible. The equations give the motion of the centre of the fictitious 'gyroscope', at the periphery of which the actual particle is supposed to be situated, so that it oscillates around the trajectory defined by the equations.

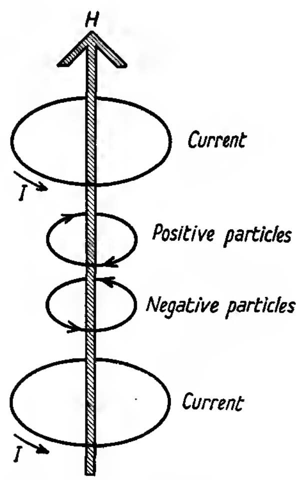

FIG. 2.1. In a magnetic field H produced by currents I, negative particles rotate in the same direction as I, positive particles in the opposite direction.

Confining ourselves to the non-relativistic case, we consider a particle with mass m and charge e moving in a homogeneous magnetic field H which is directed along the z-axis of an orthogonal coordinate system. The velocity v of the particle has the component $v_\parallel$ in the z-direction and the component $v_\perp$ in the xy-plane. If we put $W_\parallel = \frac{1}{2}mv_\parallel^2$ and $W_\perp = \frac{1}{2}mv_\perp^2$, the energy of the particle is

$$W = W_\parallel + W_\perp = \tfrac{1}{2}mv^2. \tag{1}$$

Because of the action of the magnetic field the particle moves in a spiral, the motion being composed of a motion with constant velocity in the z-direction and a circular motion in the xy-plane. In the projection of the path upon the xy-plane the radius of curvature $\boldsymbol{\rho}$ (a vector from the actual position of the particle) is given by

$$m\omega^2\boldsymbol{\rho} = (e/c)[\mathbf{vH}], \tag{2}$$

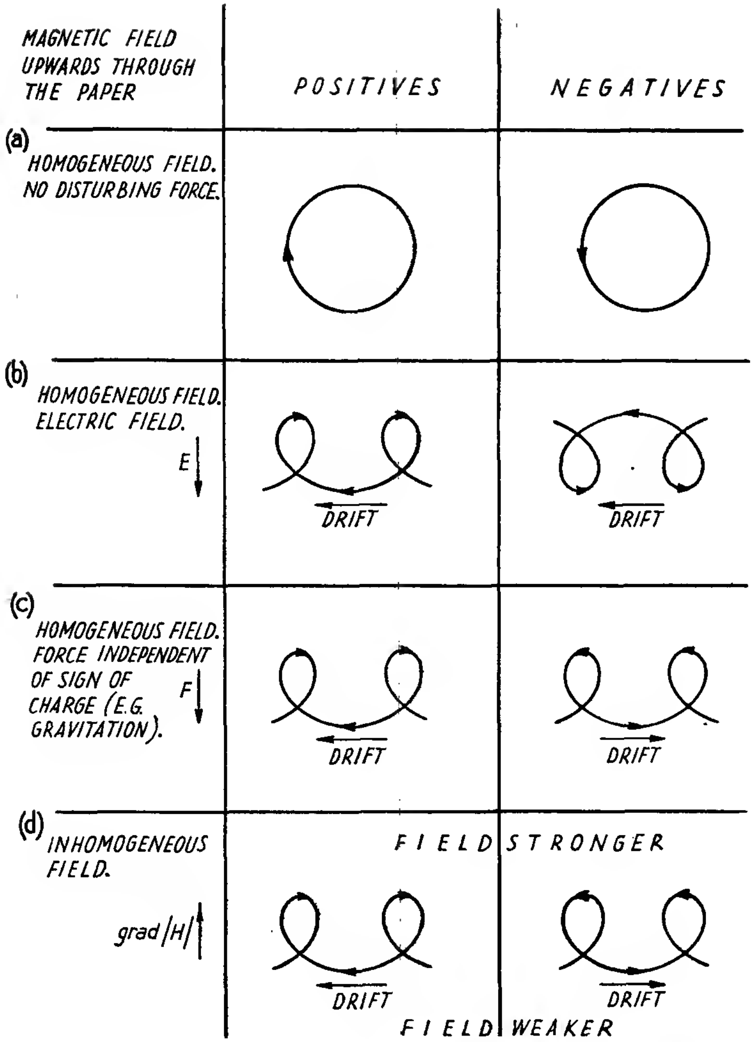

FIG. 2.2. Drifts of charged particles in a magnetic field. If the magnetic field is inhomogeneous, the radius of curvature is smaller where the field is strong than where it is weak. Hence the circle in which the particle moves in the case of a homogeneous field is changed into the curve shown above. When the inhomogeneity is small the curve is a trochoid and the motion consists of a circular motion superimposed by a translational motion ('drift') perpendicular to the magnetic gradient. [Cf. equation 2.33 (25).] An electric field E, or another force F, also changes the curvature with the results shown above. [Cf. equation 2.2 (24).]

or, as $|[\mathbf{vH}]| = Hv_\perp$,

$$\rho = \frac{cmv_\perp}{eH} = \frac{c}{eH}\sqrt{(2mW_\perp)}, \tag{2'}$$

with the gyrofrequency $$\omega = \frac{2\pi}{T} = \frac{eH}{mc}, \tag{3}$$

where T is the period (time to complete one turn):

$$T = 2\pi\frac{c}{e}\frac{m}{H}. \tag{3'}$$

The time-average of the magnetic field produced by a particle which

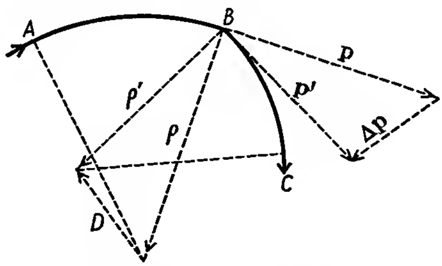

FIG. 2.3. A change in momentum $\mathbf{\Delta p}$ produces a displacement $\mathbf{D}$ of the centre of curvature.

makes a large number of loops is equivalent to a circular current

$$I = \frac{e}{c}\frac{1}{T}. \tag{4}$$

Hence the particle is equivalent to a magnet with magnetic moment $\mu = \pi\rho^2 I$, or because of (2′), (3), (3′), and (4)

$$\mu = W_\perp/H. \tag{5}$$

The magnet is antiparallel to H. The magnetic flux ($\phi = \pi\rho^2 H$) through the circular path is easily found to be

$$\phi = \frac{2\pi mc^2}{e^2}\mu. \tag{5'}$$

Introducing (3) into (2) we obtain

$$\boldsymbol{\rho} = \frac{c}{eH^2}[\mathbf{pH}], \tag{6}$$

with $$\mathbf{p} = m\mathbf{v}. \tag{6'}$$

Suppose that during a small interval of time Δt the force f with components $f_\parallel$ and $f_\perp$ acts upon the particle. Then its momentum component $p_\perp$ changes to

$$\mathbf{p}'_\perp = \mathbf{p}_\perp + \Delta\mathbf{p}_\perp, \tag{7}$$

where $$\Delta\mathbf{p}_\perp = \int_0^{\Delta t} \mathbf{f}_\perp\,dt. \tag{8}$$

Because of this change of momentum, the *centre of curvature* of the projection on the xy-plane of the path is displaced a distance

$$\mathbf{D} = \boldsymbol{\rho}' - \boldsymbol{\rho} \tag{9}$$

or, because of (6)–(8),

$$\mathbf{D} = -\frac{c}{eH^2}[\mathbf{H}\Delta\mathbf{p}_\perp], \tag{10}$$

$$\mathbf{D} = -\frac{c}{eH^2}\left[\mathbf{H}\int \mathbf{f}_\perp\, dt\right]. \tag{11}$$

This formula holds for a single collision, but of course also for a series of collisions. If $\mathbf{f}$ is a continuous force, $\mathbf{D}$ gives the displacement of that point where the centre of curvature would be if $\mathbf{f}$ vanished for a moment. This point shall be called the *guiding centre*. If $\mathbf{f}$ is continuous, *the guiding centre drifts with the velocity*

$$\mathbf{U}_\perp = \frac{d\mathbf{D}}{dt} = -\frac{c}{eH^2}[\mathbf{H}\mathbf{f}]. \tag{12}$$

When $\mathbf{f}$ has a component $f_\parallel$ in the direction of the magnetic field, an accelerated motion in this direction is obtained at the same time:

$$\frac{dU_\parallel}{dt} = \frac{1}{m}f_\parallel = \frac{1}{mH}(\mathbf{H}\mathbf{f}). \tag{13}$$

The case when $\mathbf{f}$ is continuous can also be treated in the following way. The force $\mathbf{f}_\perp$ may be due in part to an electric field $\mathbf{E}_\perp$, and in part to other forces $\mathbf{f}_\perp^0$, so that we have

$$\mathbf{f}_\perp = \mathbf{f}_\perp^0 + e\mathbf{E}_\perp. \tag{14}$$

We make a transformation to a system moving with the velocity $\mathbf{u}_\perp$ by means of the formulae (1) and (2) of § 1.3. Supposing that

$$u_\perp \ll c, \tag{15}$$

the square root, which is a relativistic correction, can be put equal to unity. Then we have

$$\mathbf{E}'_\perp = \mathbf{E}_\perp + (1/c)[\mathbf{u}_\perp \mathbf{H}], \tag{16}$$

$$\mathbf{H}' = \mathbf{H} - (1/c)[\mathbf{u}_\perp \mathbf{E}_\perp]. \tag{17}$$

When the moving system is accelerated we must introduce the inertia force

$$\mathbf{f}_\perp^i = -m\frac{d\mathbf{u}_\perp}{dt}. \tag{18}$$

This force is usually small and may often be neglected. Consequently, in the moving system the particle is acted upon by the magnetic field

$$\mathbf{H}' = \mathbf{H} - \left[\frac{\mathbf{u}_\perp}{c}\mathbf{E}_\perp\right], \tag{19}$$

and by the force $$\mathbf{f}'_\perp = \mathbf{f}^0_\perp - m\frac{d\mathbf{u}_\perp}{dt} + e\mathbf{E}'_\perp, \tag{20}$$

or, because of (14) and (16),

$$\mathbf{f}'_\perp = \mathbf{f}_\perp + \frac{e}{c}[\mathbf{u}_\perp \mathbf{H}] - m\frac{d\mathbf{u}_\perp}{dt}.$$

Putting $$\mathbf{u}_\perp = -\frac{c}{eH^2}\left[\mathbf{H}\left(\mathbf{f}_\perp - m\frac{d\mathbf{u}_\perp}{dt}\right)\right], \tag{21}$$

and observing that for a vector **A**, which is perpendicular to **H**, we have $[\mathbf{H}[\mathbf{AH}]] = H^2\mathbf{A}$, we obtain

$$\mathbf{f}'_\perp = 0. \tag{22}$$

As in the moving system no force, except the force from the magnetic field, acts in the xy-plane, the particle moves in a circle relative to the moving system. The magnetic field H' is very close to H because of (15).

Consequently when a particle moves in a homogeneous field H under the action of a force f with the components $f_\parallel$ and $f_\perp$, its path is a circle, the centre of which drifts with a velocity given by the differential equation

$$\mathbf{u}_\perp = -\frac{c}{eH^2}[\mathbf{H}\{\mathbf{f}+\mathbf{f}^i\}], \tag{23}$$

where **f** is the force applied to the particle and

$$\mathbf{f}^i = -m\frac{d\mathbf{u}}{dt}. \tag{24}$$

For negative particles the sign in (23) is positive because e is negative. In many important cases f^i is small and can be neglected in (23). The motion parallel to the field is determined by

$$(\mathbf{H}\{\mathbf{f}+\mathbf{f}^i\}) = 0, \tag{25}$$

which is equivalent to $$m\frac{d\mathbf{u}_\parallel}{dt} = \mathbf{f}_\parallel.$$

When high-order derivatives of **f** are small, the equation (23) is satisfied by the series:

$$u_x = \frac{c}{eH}\left[f_y - \left(\frac{\tau}{2\pi}\right)^2\frac{d^2f_y}{dt^2} + \left(\frac{\tau}{2\pi}\right)^4\frac{d^4f_y}{dt^4} - \dots + \frac{\tau}{2\pi}\frac{df_x}{dt} - \left(\frac{\tau}{2\pi}\right)^3\frac{d^3f_x}{dt^3} + \dots\right], \tag{26}$$

$$u_y = -\frac{c}{eH}\left[f_x - \left(\frac{\tau}{2\pi}\right)^2\frac{d^2f_x}{dt^2} + \left(\frac{\tau}{2\pi}\right)^4\frac{d^4f_x}{dt^4} - \dots - \frac{\tau}{2\pi}\frac{df_y}{dt} + \left(\frac{\tau}{2\pi}\right)^3\frac{d^3f_y}{dt^3} - \dots\right]. \tag{27}$$

It must be observed that even when $df/dt = 0$, f may depend implicitly on time.

When f varies slowly, we need often only take account of the terms containing f_x, f_y, eventually also of df_x/dt and df_y/dt.

Let us treat the motion in the xy-plane, neglecting all terms except f_x, f_y. Then $U_\perp$ (from (12)) and $u_\perp$ are equal. If **f** derives from a potential, the drift, which is perpendicular to **f**, follows *an equipotential line*. The average energy of the particle is constant.

If in the moving system the circular velocity is $v_\perp$, and consequently the kinetic energy $W_\perp = \frac{1}{2}mv_\perp^2$, the energy in the fixed system is

$$W'_\perp = \tfrac{1}{2}mv'^2_\perp = \tfrac{1}{2}m(\mathbf{v}_\perp+\mathbf{u}_\perp)^2 = W_\perp+\tfrac{1}{2}mu_\perp^2+m(\mathbf{u}_\perp\mathbf{v}_\perp). \tag{28}$$

Averaging over one turn, the scalar product cancels, so that we obtain for the mean values $\overline{W}_\perp$ and $\overline{W}'_\perp$

$$\overline{W}'_\perp = \overline{W}_\perp+\tfrac{1}{2}mu_\perp^2. \tag{29}$$

When the terms $df_\perp/dt$ are also taken into account, there is a difference between $U_\perp$ and $u_\perp$. The 'guiding centre' (to which $U_\perp$ refers) moves always perpendicular to f, but the centre of the circle in the moving system (to which $u_\perp$ refers) has also a velocity component in the direction of $df_\perp/dt$. Hence the average energy $\overline{W}'$ is changed at the rate

$$\frac{d\overline{W}'_\perp}{dt} = (\mathbf{f}\mathbf{u}_\perp) = -\frac{c}{eH^2}(\mathbf{f}[\mathbf{H}(\mathbf{f}+\mathbf{f}^i)]) = \frac{c}{eH}(\mathbf{f}^i[\mathbf{H}(\mathbf{f}+\mathbf{f}^i)])$$

$$= m\mathbf{u}_\perp\frac{d\mathbf{u}_\perp}{dt} = \frac{d}{dt}\left(\frac{mu_\perp^2}{2}\right). \tag{30}$$

Hence the centre of the circle may be displaced to another equipotential line by a change in **f**. The difference in energy between the two equipotential lines is equal to the change in kinetic energy due to the change in drift velocity. The circling velocity, i.e. the velocity in relation to the moving system, remains constant.

2.3. Inhomogeneous magnetic field

The equations (23) and (25) of § 2.2 may be generalized to the case when the magnetic field is a function of space and time with the condition that the change during one turn is small. We assume that

$$|(\boldsymbol{\rho}\,\mathrm{grad})\mathbf{H}| \ll |H| \tag{1}$$

and

$$T\left|\frac{d\mathbf{H}}{dt}\right| \ll |H|. \tag{2}$$

We introduce the effects of the inhomogeneity, and of time variation, of the magnetic field as perturbations of the motion in the homogeneous field. We have to consider perturbations of three different kinds:

2.31. *The magnetic field varies with time.* Then an electromotive

force is induced, which changes the energy of the particle. We have

$$\oint (\mathbf{E}\,d\mathbf{s}) = -\frac{1}{c}\frac{d\phi}{dt}, \tag{3}$$

where ϕ $(= \pi\rho^2 H)$ is the flux through the circular path of the particle and the integral is to be taken along the periphery of the same circle. The gain in energy during one turn is

$$\Delta W_\perp = -|e| \oint \mathbf{E}\,d\mathbf{s} = \frac{\pi\rho^2|e|}{c}\frac{dH}{dt}. \tag{4}$$

(The negative sign derives from the fact that a positive particle goes in a direction opposite to that in which the integral is to be taken.) Thus the rate of increase in energy is given by

$$\frac{dW_\perp}{dt} = \frac{\Delta W_\perp}{T} = \frac{W_\perp}{H}\frac{dH}{dt}, \tag{5}$$

using equations (2′) and (3′) of § 2.2. This shows that the magnetic moment μ $(= W_\perp/H)$ remains constant when the magnetic field changes.

The general electric field produced by the changing magnetic field has to be introduced as in § 2.2.

2.32. *The gradient of the magnetic field has a component in the direction of the field,* so that if at a point we place an orthogonal reference system with the z-axis parallel to the magnetic field we have $dH/dz \neq 0$. Introducing polar coordinates (ρ, ϑ) in the xy-plane, the condition

$$\operatorname{div}\mathbf{H} = 0 \tag{6}$$

gives (if $H_\vartheta = 0$)

$$\frac{1}{\rho}\frac{d}{d\rho}(\rho H_\rho) + \frac{dH_z}{dz} = 0, \tag{7}$$

where H_z, H_ρ, and H_ϑ are the components of $\mathbf{H}$. Because of (1) we can put $dH_z/dz = dH/dz =$ constant over the small circle ρ so that we obtain

$$H_\rho = -\tfrac{1}{2}\rho\frac{dH}{dz}. \tag{8}$$

As the particle moves with the velocity $v_\perp$ in the xy-plane it is subject to a force

$$f_z^{(m)} = (e/c)v_\perp H_\rho \tag{9}$$

in the direction of the z-axis. Introducing (8), equations (2′) and (5) from § 2.2, and $v_\perp = \sqrt{(2W_\perp/m)}$, we obtain

$$f_z^{(m)} = -\mu\frac{dH}{dz}. \tag{10}$$

It is immaterial whether H represents H_z or $\sqrt{(H_x^2+H_y^2+H_z^2)}$.

When the particle moves in the direction of the z-axis a distance Δz, the energy component $W_{\parallel}$ is increased

$$\Delta W_{\parallel} = f_z^{(m)} \Delta z. \tag{11}$$

As the total energy W $(= W_{\parallel}+W_{\perp})$ remains constant we have

$$\Delta W_{\perp} = -f_z^{(m)} \Delta z. \tag{12}$$

Owing to the displacement of the particle there is a change in the magnetic field in which the particle moves:

$$\Delta H = \frac{dH}{dz} \Delta z. \tag{13}$$

Introducing equations (10) and 2.2 (5) into (12) and eliminating $(dH/dz)\,\Delta z$ by the help of (13), we find

$$\Delta W_{\perp} = \frac{W_{\perp}}{H} \Delta H, \tag{14}$$

showing that even in this case $\mu = W_{\perp}/H$ remains constant. According to equation 2.2 (5′) this means that the flux through the circular path remains constant. The particle moves on the surface of a flux tube.

2.33. *The gradient of the magnetic field has a component perpendicular to the magnetic field.* Suppose that in the xy-plane (which is perpendicular to H) $dH/dx = 0$ but $dH/dy = b \neq 0$. If H_0 is the magnetic field at the centre of the circular path of a positive particle, which at time $t = 0$ is at the point $(\rho, 0)$, then at the time t the particle is situated in a field of which the strength

$$H = H_0 - b\rho \sin \omega t, \tag{15}$$

where ρ and ω are given by equations 2.2 (2′) and 2.2 (3). The components of momentum of the particle are

$$p_x = -p_{\perp} \sin \omega t \tag{16}$$

$$p_y = -p_{\perp} \cos \omega t. \tag{17}$$

When H is increased by dH, the radius of curvature changes according to equation 2.2 (6). Hence the centre of curvature is displaced a distance $\mathbf{ds}$ in the direction of $\boldsymbol{\rho}$

$$\mathbf{ds} = -\frac{c\,dH}{eH^3}[\mathbf{pH}]. \tag{18}$$

Introducing (16), (17), and, after differentiation, (15), and observing that $b\rho \ll H_0$, we obtain for the components of ds

$$\frac{ds_x}{dt} = -\frac{c}{eH^2} p_{\perp}\, b\rho\omega \cos^2 \omega t, \tag{19}$$

$$\frac{ds_y}{dt} = \frac{c}{eH^2} p_{\perp}\, b\rho\omega \sin \omega t \cos \omega t. \tag{20}$$

The drift $u_\perp$ (with components u_x and u_y) is the average value of (19) and (20) and can be calculated from

$$u_x = \frac{1}{T}\int\limits_0^T ds_x \tag{21}$$

and a corresponding formula for u_y. After elementary reductions we obtain

$$u_x = -\frac{c}{eH}\mu\frac{dH}{dy}, \qquad u_y = 0, \tag{22}$$

where μ is the magnetic moment:

$$\mu = \frac{W_\perp}{H}. \tag{23}$$

$W_\perp$ is the energy of the circling velocity, i.e. the kinetic energy referred to a system where the centre is at rest.

Introducing $$\mathbf{f}^{(m)} = -\mu \operatorname{grad} H, \tag{24}$$

we obtain $$\mathbf{u}_\perp = -\frac{c}{eH^2}[\mathbf{H}\mathbf{f}^{(m)}]. \tag{25}$$

For a particle which drifts perpendicularly to the magnetic field under the influence of a force (e.g. an electric field) and an inhomogeneity of the magnetic field, we have to add to (25) the drift found from equation 2.2 (23). The displacement of the particle changes the field in which it moves at the rate

$$\frac{dH}{dt} = \mathbf{u}_\perp \operatorname{grad} H = -\frac{c}{eH^2}[\mathbf{H}\{\mathbf{f}+\mathbf{f}^m+\mathbf{f}^i\}]\operatorname{grad} H. \tag{26}$$

Here $\mathbf{f}^m = -\mu \operatorname{grad} H$. For an arbitrary vector $\mathbf{A}$, we have $([\mathbf{HA}]\mathbf{A}) = 0$. Hence

$$\frac{dH}{dt} = -\frac{c}{eH^2}[\mathbf{H}\{\mathbf{f}+\mathbf{f}^i\}]\operatorname{grad} H. \tag{27}$$

At the same time the average energy changes. As the force from the magnetic field acts perpendicularly to v, it produces no change in energy. Thus we have

$$\frac{dW'_\perp}{dt} = (\mathbf{u}_\perp \mathbf{f}). \tag{28}$$

The average energy referred to the moving system, according to equation 2.2 (29), is given by

$$W_\perp = W'_\perp - \tfrac{1}{2}mu_\perp^2.$$

Differentiating this we obtain from (28)

$$\frac{dW_\perp}{dt} = \left(\mathbf{u}_\perp\left\{\mathbf{f}-m\frac{d\mathbf{u}_\perp}{dt}\right\}\right) = (\mathbf{u}_\perp\{\mathbf{f}+\mathbf{f}^i\}) = -\frac{c}{eH^2}([\mathbf{H}\{\mathbf{f}+\mathbf{f}^m+\mathbf{f}^i\}]\{\mathbf{f}+\mathbf{f}^i\}) \tag{29}$$

or, as

$$([\mathbf{H}\{\mathbf{A}_1+\mathbf{A}_2\}]\mathbf{A}_1) = -([\mathbf{H}\mathbf{A}_1]\mathbf{A}_2),$$

$$\frac{dW_\perp}{dt} = \frac{c}{eH^2}[\mathbf{H}(\mathbf{f}+\mathbf{f}^i)]\mathbf{f}^m. \tag{30}$$

As $\mathbf{f}^m = -\mu\,\mathrm{grad}\,H = -(W_\perp/H)\mathrm{grad}\,H$, we have from (27) and (30)

$$\frac{dW_\perp}{dH} = \frac{W_\perp}{H}, \tag{31}$$

showing that μ remains constant.

As the perturbation method is applicable only when the inhomogeneity is small, this magnetic drift is small in comparison to the circling velocity v of the particle. The drift due to the electric field, however, is not subject to the same restriction. It could very well be larger than v, but it must always remain small compared with the velocity of light.

Hence an inertia force f^i due to a change in the electric field may be important, but a change in the magnetic inhomogeneity does not produce a considerable inertia force (within the limits of applicability of the perturbation method).

The result is that a particle in a magnetic field, the inhomogeneity of which is small ($\rho|\mathrm{grad}\,H| \ll |H|$), moves in a circle, the centre of which drifts with a velocity u, which has a component perpendicular to the field:

$$\mathbf{u}_\perp = -\frac{c}{eH^2}[\mathbf{H}(\mathbf{f}+\mathbf{f}^m+\mathbf{f}^i)], \tag{32}$$

where $\mathbf{f}$ is the sum of non-magnetic forces applied to the particle, and

$$\mathbf{f}^m = -\mu\,\mathrm{grad}\,H, \tag{33}$$

$$\mathbf{f}^i = -m\frac{d\mathbf{u}}{dt}. \tag{34}$$

The velocity component $u_\parallel$ parallel to the field is given by (34) combined with

$$(\mathbf{H}\{\mathbf{f}+\mathbf{f}^m+\mathbf{f}^i\}) = 0. \tag{35}$$

During the motion μ remains constant.

The force f may be due to an electric field or to gravitation. It may also be composed of a series of impacts (compare § 2.2), as is the case when the particle is part of a gas which is subject to a pressure gradient. In this case the force is given by

$$\mathbf{f} = -(1/n)\mathrm{grad}\,p, \tag{36}$$

where n is the number of particles per unit volume and p the pressure.

2.4. Motion in a dipole field. Störmer's method

The general case of the motion of a charged particle in a dipole field has been treated by Störmer. Only a brief account will be given here. For a detailed study the reader is referred to Störmer's original works or the surveys included in most text-books.

If the magnetic field from a dipole with moment a is H, and e, m, and $\mathbf{v}$ are the charge, mass, and velocity of the particle, the equation of motion is

$$m\frac{d\mathbf{v}}{dt} = \frac{e}{c}[\mathbf{vH}]. \tag{1}$$

Introducing cylindrical coordinates (r, λ, z), the z-axis being parallel to the dipole, we have for the components of (1)

$$\frac{d^2r}{dt^2}-r\left(\frac{d\lambda}{dt}\right)^2 = \frac{e}{mc}H_z r\frac{d\lambda}{dt}, \tag{2}$$

$$\frac{d^2z}{dt^2} = -\frac{e}{mc}H_r r\frac{d\lambda}{dt}, \tag{3}$$

$$\frac{d}{dt}\left(r^2\frac{d\lambda}{dt}\right) = \frac{er}{mc}\left(H_r\frac{dz}{dt}-H_z\frac{dr}{dt}\right). \tag{4}$$

We put $ds = \mathrm{v}\, dt$ and introduce the length unit c_{St},

$$c_{\mathrm{St}} = \sqrt{\left(\frac{|e|a}{mcv}\right)} = \sqrt{\left(\frac{|e|a}{cp}\right)} = \sqrt{\left(\frac{a}{H\rho}\right)}, \tag{5}$$

where $p = mv$ is the momentum. We shall discuss the case of a positive particle. For negative particles the paths are the mirror images, with respect to a plane through the z-axis, of the paths of positive particles.

For the purely mathematical discussion it is convenient to express all lengths in c_{St} as unit, but for the physical application it is preferable to have this quantity included in the formulae.

The velocity v is constant because the force is always perpendicular to the velocity. Using equations (12)–(14) of § 1.2, equation (4) gives for a positive charge

$$\frac{d}{ds}\left(r^2\frac{d\lambda}{ds}\right) = c_{\mathrm{St}}^2\left(\frac{3r^2z}{R^5}\frac{dz}{ds}+\frac{R^2-3z^2}{R^5}r\frac{dr}{ds}\right)$$

$$= -c_{\mathrm{St}}^2\left[\frac{\partial}{\partial z}\left(\frac{r^2}{R^3}\right)\frac{dz}{ds}+\frac{\partial}{\partial r}\left(\frac{r^2}{R^3}\right)\frac{dr}{ds}\right]$$

with $R = \sqrt{(z^2+r^2)}$. After integration we obtain

$$\frac{r^2}{c_{\mathrm{St}}}\frac{d\lambda}{ds} = -c_{\mathrm{St}}\frac{r^2}{R^3}-2\gamma, \tag{6}$$

where 2γ is an integration constant proportional to the angular momentum at infinity. Observing that

$$r^2\left(\frac{d\lambda}{ds}\right)^2+\left(\frac{dr}{ds}\right)^2+\left(\frac{dz}{ds}\right)^2=1,$$

we obtain from (2), (3), and (6)

$$\frac{1}{c_{\mathrm{St}}^4}\frac{d^2r}{ds^2}=\left(\frac{2\gamma}{c_{\mathrm{St}}\,r}+\frac{r}{R^3}\right)\left(\frac{2\gamma}{c_{\mathrm{St}}\,r^2}+\frac{3r^2}{R^5}-\frac{1}{R^3}\right), \tag{7}$$

$$\frac{1}{c_{\mathrm{St}}^4}\frac{d^2z}{ds^2}=\left(\frac{2\gamma}{c_{\mathrm{St}}\,r}+\frac{r}{R^3}\right)\frac{3rz}{R^5}, \tag{8}$$

$$\left(\frac{dr}{ds}\right)^2+\left(\frac{dz}{ds}\right)^2=1-\left(\frac{2\gamma}{c_{\mathrm{St}}\,r}+\frac{r}{R^3}\right)^2c_{\mathrm{St}}^4. \tag{9}$$

Putting the velocity component $r\,d\lambda/dt$ $(= rv\,d\lambda/ds)$ equal to $v\sin\theta$, we have

$$\sin\theta=r\frac{d\lambda}{ds}=-\left(\frac{c_{\mathrm{St}}^2\,r}{R^3}+\frac{2\gamma c_{\mathrm{St}}}{r}\right). \tag{10}$$

As $|\sin\theta|\leqslant 1$, we obtain

$$\left|2\gamma\frac{c_{\mathrm{St}}}{r}+\frac{r}{R}\frac{c_{\mathrm{St}}^2}{R^2}\right|\leqslant 1. \tag{11}$$

Störmer has shown that the value of γ determines the character of the orbits. If $\gamma>0$ the orbits never reach the dipole. For $-1\leqslant\gamma\leqslant 0$, particles can move up to the dipole from infinity. One of the orbits for $\gamma=-1$ is a circle in the equatorial plane with $R=c_{\mathrm{St}}$. If $\gamma<-1$, there are two different allowed regions, one outer region ($R>c_{\mathrm{St}}$) far away from the dipole and one inner region ($R<c_{\mathrm{St}}$) close to the dipole. The latter group contains those periodic (or quasiperiodic) orbits which may be treated by the perturbation method.

The following table gives the value of c_{St} in some cases of interest:

TABLE 2.1

$H\rho$ *gauss.cm.*	*Electron-volts*		$c_{\mathrm{St}}=\sqrt{(a/H\rho)}$ *cm.*	
	Electron	*Proton*	*Sun* $a=4{\cdot}2.10^{33}$	*Earth* $a=8{\cdot}2.10^{25}$
10^2	$0{\cdot}89.10^3$	$0{\cdot}48$	$6{\cdot}5.10^{15}$	$9{\cdot}1.10^{11}$
10^3	$0{\cdot}82.10^5$	48	$2{\cdot}0.10^{15}$	$2{\cdot}9.10^{11}$
10^4	$2{\cdot}53.10^6$	$4{\cdot}8.10^3$	$6{\cdot}5.10^{14}$	$9{\cdot}1.10^{10}$
10^5	$2{\cdot}95.10^7$	$4{\cdot}8.10^5$	$2{\cdot}0.10^{14}$	$2{\cdot}9.10^{10}$
10^6	$3{\cdot}0.10^8$	$4{\cdot}7.10^7$	$6{\cdot}5.10^{13}$	$9{\cdot}1.10^9$
10^7	$3{\cdot}0.10^9$	$2{\cdot}2.10^9$	$2{\cdot}0.10^{13}$	$2{\cdot}9.10^9$
10^8	$3{\cdot}0.10^{10}$	$2{\cdot}9.10^{10}$	$6{\cdot}5.10^{12}$	$9{\cdot}1.10^8$
10^9	$3{\cdot}0.10^{11}$	$3{\cdot}0.10^{11}$	$2{\cdot}0.10^{12}$	$2{\cdot}9.10^8$

Compare: earth's radius $=6{\cdot}37.10^8$ cm., sun's radius $=7{\cdot}0.10^{10}$ cm., distance earth–moon $=3{\cdot}8.10^{10}$ cm., distance sun–earth $=1{\cdot}49.10^{13}$ cm.; solar magnetic field equals terrestrial field at 4.10^{10} cm. from the earth.

Except in problems concerning cosmic rays, there is no problem relating to the terrestrial field where the electronic energy exceeds 10^5 volts. This corresponds to $c_{St} = 2{\cdot}5.10^{11}$ cm. In the solar field the same energy gives $c_{St} = 1{\cdot}7.10^{15}$ cm. Consequently in all problems, except cosmic rays, we have $r \leqslant R \ll c_{St}$. Excluding the exceptional case $r \ll R$ (which corresponds to motions very close to the magnetic axis), we find from (11) that we must have

$$\gamma \ll -1. \tag{12}$$

In this case, which consequently is of most interest in cosmic physics, the particle moves in orbits whose radii of curvature are small in comparison to the distance from the dipole. The perturbation method is well suited to the treatment of this problem (see § 2.5).

The conditions in the equatorial plane ($r = R$) are simple. Particles reach the equator at an angle θ given by the equation (10)

$$\sin\theta = -\frac{c_{St}}{R_0}\left(2\gamma + \frac{c_{St}}{R_0}\right). \tag{13}$$

If the momentum is given, c_{St} is defined. For $c_{St} > R_0$ and for a certain value of θ we have an orbit coming from infinity if $\gamma > -1$, and a periodic orbit (never leaving the neighbourhood of the earth) if $\gamma < -1$.

The boundary between orbits from infinity and periodical orbits is given by $\gamma = -1$. For $\sin\theta = +1$ we have

$$R_0 = c_{St}.$$

According to (5) this corresponds to a momentum

$$p_2 = \frac{ea}{cR_0^2}. \tag{14}$$

Particles above this momentum can reach a point at the equator from all directions. For $\sin\theta = -1$ we obtain

$$R_0(1+\sqrt{2}) = c_{St}.$$

The corresponding momentum is

$$p_1 = \frac{ea}{cR_0^2}(3-2\sqrt{2}). \tag{15}$$

Particles below this momentum cannot reach the equator at all. For p-values in the range $p_1 < p < p_2$, particles are allowed within a cone defined by θ according to (13).

For higher latitudes particles are also allowed within certain cones, but these are usually very complex. Much labour has been devoted to

these problems which are very important in the study of cosmic radiation. Besides Störmer, Lemaître and Vallarta and many others have made extensive investigations in this field. A detailed account of their results is beyond the scope of this book. Recently a summary has been given by Meixner (1943).

2.5. Motion in a magnetic dipole field; perturbation method

In most cases of interest—in fact all problems except those concerning cosmic rays—the radius of curvature is small compared with the distance from the dipole, which means that we can use the perturbation method. The calculations may be carried out in the following way.

We start from § 2.3, equations (32) and (35), where f equals zero. In order to calculate f^m we employ polar coordinates: radius vector $= R$, magnetic latitude $= \varphi$, magnetic longitude $= \lambda$. For this system the equations of the magnetic field are given in § 1.2.

The magnetic energy ϵ of a magnet having the moment μ and situated antiparallel to the field H is given by

$$\epsilon = -\mu H = -\frac{\mu a}{R^3}\phi. \tag{1}$$

The force acting upon this dipole has the components

$$f_R = \frac{\delta\epsilon}{\delta R} = \frac{3\mu a}{R^4}\phi, \tag{2}$$

$$f_\varphi = \frac{\delta\epsilon}{R\,\delta\varphi} = -\frac{3\mu a}{R^4}\frac{\sin\varphi\cos\varphi}{\phi}, \tag{3}$$

$$f_\lambda = 0.$$

With the help of equations 1.2 (8), 1.2 (9), and 2.2 (5) we can now calculate the forces parallel to and perpendicular to the magnetic field.

$$f_{\parallel}^{(m)} = f_R\cos\alpha - f_\varphi\sin\alpha = \frac{3W_\perp}{R}\frac{\sin\varphi(3+5\sin^2\varphi)}{\phi^3}, \tag{4}$$

$$f_{\perp}^{(m)} = f_R\sin\alpha + f_\varphi\cos\alpha = \frac{3W_\perp}{R}\frac{\cos\varphi(1+\sin^2\varphi)}{\phi^3}. \tag{5}$$

The force $f_\perp$ (as well as $f_\parallel$) is situated in the $R\varphi$-plane.

In order to calculate $u_\perp$ from equation 2.3 (32) we must also find $f_\perp^i$. The only considerable component of this force is the centrifugal force f_c deriving from the motion with velocity $u_\parallel$ along the curved magnetic lines of force. If R_c is the radius of curvature of the line of force we have

$$f_c = mv_\parallel^2/R_c. \tag{6}$$

As a simple geometrical consideration shows, we have

$$f_{\perp}^{m} = \frac{\mu H}{R_c} = \frac{W_{\perp}}{H}\frac{H}{R_c}, \tag{7}$$

which gives

$$\mathbf{f}_c = \mathbf{f}_{\perp}^{m}\frac{2W_{\parallel}}{W_{\perp}}. \tag{8}$$

We introduce an orthogonal coordinate system with the z-axis parallel to H and the y-axis parallel to $f_{\perp}^{(m)}$. Then the equations of motion are:

$$R\cos\varphi\frac{d\lambda}{dt} = u_x = \frac{c}{eH}(f_{\perp}^{m}+f_c), \tag{9}$$

$$u_y = 0, \tag{10}$$

$$\frac{du_z}{dt} = \frac{f_{\parallel}^{(m)}}{m}. \tag{11}$$

The integration of (11) gives the same result as can be obtained immediately from equations 2.2 (1) and 2.2 (5).

$$u_z = v_{\parallel} = \sqrt{\{2(W-W_{\perp})/m\}} = \sqrt{\{2\mu(H_0-H)/m\}}, \tag{12}$$

where H_0 $(= W/\mu)$ is a constant denoting the strength of the field at the turning-point. From equation 1.2 (6) we find

$$v_{\parallel} = \frac{\sqrt{(dR^2+R^2\,d\varphi^2)}}{dt} = r_e\cos\varphi\sqrt{(1+3\sin^2\varphi)}\frac{d\varphi}{dt}. \tag{13}$$

Putting

$$H = a\eta/r_e^3; \qquad H_0 = a\eta_0/r_e^3 \tag{14}$$

in analogy with equation 1.2 (10), we obtain

$$\frac{d\varphi}{dt} = \left(\frac{2\mu a}{mr_e^5}\frac{\eta_0-\eta}{\cos^2\varphi(1+3\sin^2\varphi)}\right)^{\frac{1}{2}}. \tag{15}$$

Further, combining equations (9), (8), and (5) of this section with (5) of § 2.2, and (6) and (14) of § 1.2, we find

$$\frac{d\lambda}{dt} = \frac{3c\mu}{er_e^2}\frac{1+\sin^2\varphi}{\cos^4\varphi(1+3\sin^2\varphi)^{\frac{3}{2}}}\frac{2\eta_0-\eta}{\eta}. \tag{16}$$

We can now compute the *path of the particle*. Introducing Störmer's unit of length

$$c_{\mathrm{St}} = \sqrt{\left(\frac{ae}{cmv}\right)} = \sqrt[4]{\left(\frac{ae^2r_e^3}{2c^2m\mu\eta_0}\right)} \tag{17}$$

and equation 1.2 (10) we find

$$\lambda-\lambda_0 = (r_e/c_{\mathrm{St}})^2 I_1, \tag{18}$$

with

$$I_1 = 3\int\limits_0^{\phi}\frac{\cos^3\varphi(1+\sin^2\varphi)}{(1+3\sin^2\varphi)^{\frac{3}{2}}}\frac{1-\frac{1}{2}\eta(\varphi)/\eta_0}{1-\eta(\varphi)/\eta_0}\,d\varphi, \tag{18'}$$

where r_e is the distance from the dipole to the points where the particle (or more exactly the 'equivalent magnet') crosses the equatorial plane ($\varphi = 0$); $\eta(\varphi)$ is defined by equation 1.2 (11) and η_0 is a constant [see (14)].

The parameter r_e is related to Störmer's constant γ. For the equa-

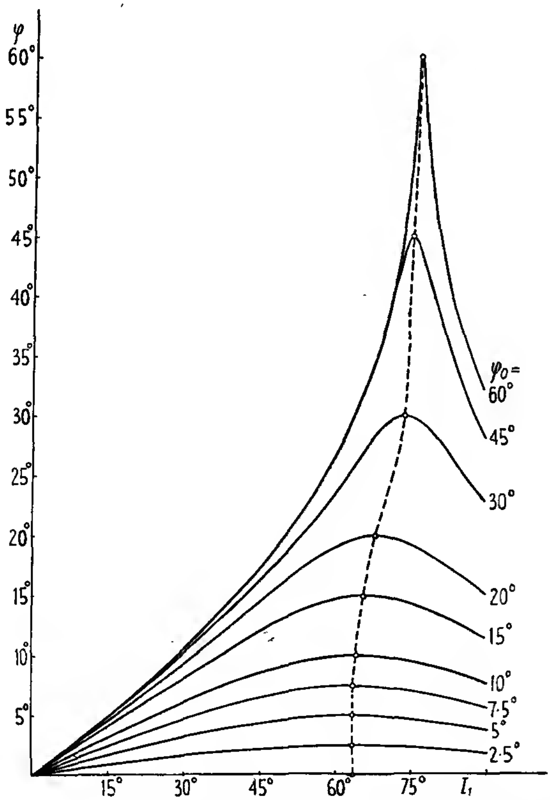

Fig. 2.4. Connexion between displacement in longitude (proportional to I_1) and latitude ϕ for a particle oscillating through the equatorial plane with amplitude φ_0.

torial plane the condition 2.4 (11) can be written

$$-1 \leqslant \frac{c_{\rm St}}{R}\left(2\gamma+\frac{c_{\rm St}}{R}\right) \leqslant +1. \tag{19}$$

As R oscillates between the limits $r_e+\rho$ and $r_e-\rho$ we have ($r_e \ll c_{\rm St}$):

$$r_e/c_{\rm St} = -\tfrac{1}{2}\gamma^{-1} \tag{20}$$

and

$$\rho/r_e = (r_e/c_{st})^2 = \tfrac{1}{4}\gamma^{-2}. \tag{20'}$$

The integral I_1 in (18′) is plotted in Fig. 2.4.

The equation (18) gives the path of the 'equivalent magnet'. The path of the particle itself is a spiral around the curve defined by (18). The spiral has the radius of curvature given by equation 2.2 (2′). In most of the cases to which our perturbation method of calculation is applicable, the equation (18) gives as much information about the motion of the particle as is wanted.

The motion defined by (15), (16), and (18) takes place on the surface defined by equation 1.2 (6), and is an oscillation through the equatorial plane $\varphi = 0$, combined with a rotation around the axis of the dipole. The amplitude of the oscillation is defined by the condition $\eta_0 - \eta \geqslant 0$. The φ-value of the turning-point is given by

$$\eta_0 = \frac{\sqrt{(1+3\sin^2\varphi_0)}}{\cos^6\varphi_0}. \tag{21}$$

Fig. 2.5 shows a comparison between one of the trajectories integrated by Störmer (1913) and the corresponding path found by the perturbation method (Alfvén 1940).

2.51. *Motion close to the equatorial plane of a dipole field.* Of particular interest is the special case when the amplitude of the oscillation is small ($\varphi_0 \ll 1$). Then we have approximately

$$\eta = \frac{\sqrt{(1+3\sin^2\varphi)}}{\cos^6\varphi} = 1+4{\cdot}5\varphi^2 \tag{22}$$

and in the same way $\eta_0 = 1+4{\cdot}5\varphi_0^2$. Putting these values into (15) and integrating, we obtain a sine oscillation

$$\varphi = \varphi_0 \sin\left(2\pi\frac{t-t_0}{T}\right), \tag{23}$$

where
$$T = \frac{2\pi}{3}\sqrt{\left(\frac{mr_e^5}{\mu a}\right)} = \frac{2\pi}{3}\sqrt{2}\frac{r_e}{v}. \tag{24}$$

To the same approximation, λ increases at a constant rate:

$$\lambda = \frac{3c\mu}{er_e^2}(t-t_1) = \frac{3}{2}\frac{r_e}{c_{\mathrm{St}}^2}v(t-t_1). \tag{25}$$

During the period T the increase in λ amounts to

$$\Lambda = \pi\sqrt{2}\, r_e^2/c_{\mathrm{St}}^2, \tag{26}$$

where c_{St} is given by equation 2.4 (5). For small amplitudes the value of $\Lambda(c_{\mathrm{St}}/r_e)^2$ is $\pi\sqrt{2}$ (= 4·44), corresponding to 255°. The dotted line in Fig. 2.4 represents the difference in longitude between the turning-point and the intersection of the orbit with the equatorial plane. Its value for $\varphi \to 0$ is $\frac{1}{4}\pi\sqrt{2}$ (= 1·11), corresponding to 63·6°.

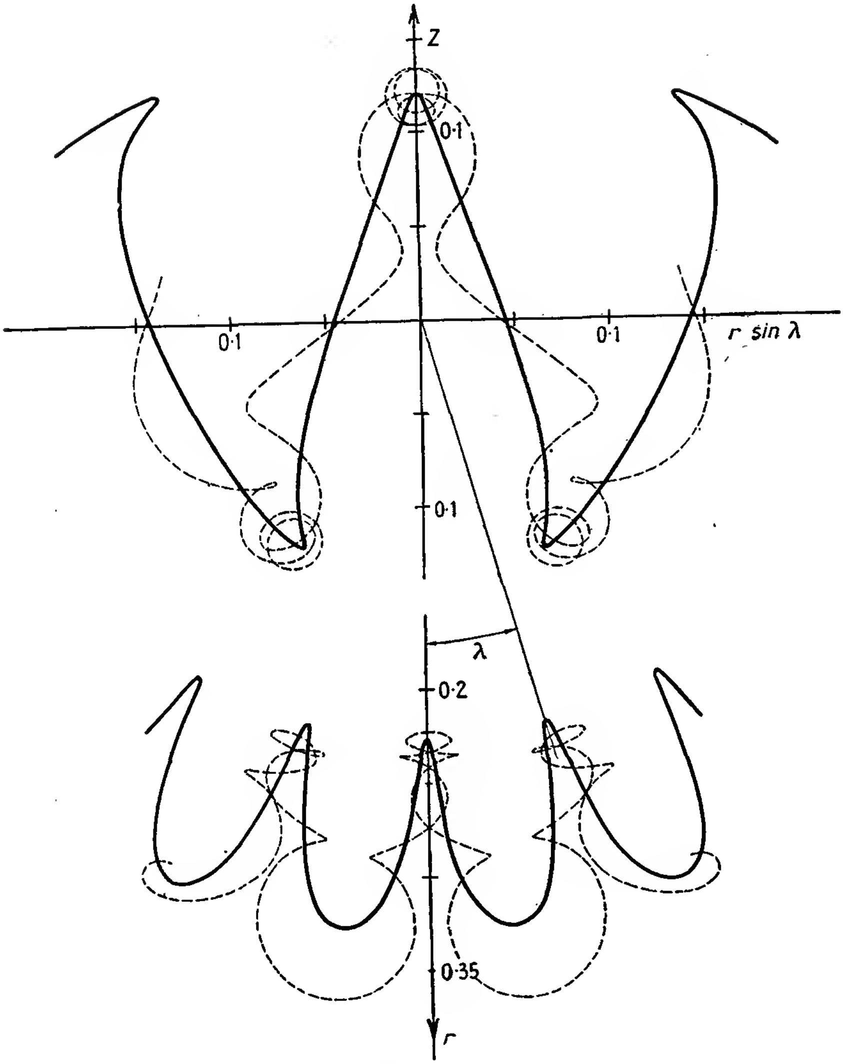

FIG. 2.5. Motion in dipole field calculated by Störmer and by perturbation method. Projection upon a plane through the axis of the dipole (above) and upon the equatorial plane (below).
——— Path of equivalent magnet. - - - - - Path of particle according to Störmer.

2.52. We now have to calculate the parameters h_0, r_e, and c_{St}. Suppose that the particle (mass $= m$) starts at the point (R', ϕ', λ') with the velocity $(v'_R, v'_\phi, v'_\lambda)$. Then we have:

$$\phi' = (1+3\sin^2\varphi')^{\frac{1}{2}}, \tag{27}$$

$$v'_{\parallel} = \frac{2v'_R \sin\varphi'_c - v'_\varphi \cos\varphi'}{\phi'}, \tag{28}$$

$$v'_{\perp} = \left\{\frac{(v'_R \cos\varphi' + 2v'_\varphi \sin\varphi')^2}{\phi'^2} + v'^2_\lambda\right\}^{\frac{1}{2}}, \tag{29}$$

$$H' = \frac{a}{R'^3}\phi'; \qquad \eta' = (\cos\varphi')^{-6}\phi', \tag{30}$$

$$\mu = \frac{mR'^3(v'_R \cos\varphi' + 2v'_\varphi \sin\varphi')^2 + \phi'^2 v'^2_\lambda}{a\phi'}, \tag{31}$$

$$r_e = R'(\cos\varphi')^{-2}, \tag{32}$$

$$\eta_0 = \eta'[1+(v'_{\parallel}/v'_{\perp})^2], \tag{33}$$

$$c_{\mathrm{St}} = \sqrt{\left(\frac{ae}{mc}\frac{1}{v'}\right)}. \tag{34}$$

The particle spirals in a circle of radius ρ which moves according to what is said above. We have

$$\rho = \frac{r_e^3}{c_{\mathrm{St}}^2\sqrt{\eta^0}}\frac{1}{\sqrt{\eta}}. \tag{35}$$

2.53. In order to show the connexion between Störmer's equations and the perturbation method, we shall derive (25) through successive approximations, valid if $\gamma \ll -1$. We put

$$r_0 = r_e/c_{\mathrm{St}} = -\tfrac{1}{2}\gamma^{-1}. \tag{36}$$

Consequently $r_0 \ll 1$.

As in the equatorial plane $z = 0$, we obtain from equation 2.4 (9)

$$\left(\frac{dr}{ds}\right)^2 = 1 - c_{\mathrm{St}}^4\frac{1}{r^4}\left(1-\frac{r}{r_e}\right)^2. \tag{37}$$

We develop r into a series

$$r = r_0 c_{\mathrm{St}}[1+r_0^2 F(s)+r_0^4 G(s)], \tag{38}$$

neglecting higher terms. Then we have (denoting d/ds by dashes)

$$dr/ds = r_0^3 c_{\mathrm{St}}(F'+r_0^2 G') \tag{39}$$

and

$$1-r/r_0 c_{\mathrm{St}} = -r_0^2(F+r_0^2 G). \tag{40}$$

The first approximation solution of (37) is obtained from

$$r_0^6 c_{\mathrm{St}}^2(F')^2 = 1-F^2, \tag{41}$$

which gives

$$F = \sin S, \tag{42}$$

where

$$S = \frac{s}{r_0^3 c_{\mathrm{St}}}. \tag{43}$$

In order to obtain the second approximation we put (42) into (38), (39), and (40) and obtain from (37)

$$r_0^6 c_{St}^2(F'^2+2r_0^2 F'G') = 1-(F^2+2r_0^2 FG)(1-4r_0^2 F). \tag{44}$$

Using (42) we get

$$G = 4-2\sin^2 S. \tag{45}$$

Consequently, equation (38) gives

$$r = r_0 c_{St}[1+r_0^2 \sin S+r_0^4(4-2\sin^2 S)], \tag{46}$$

where terms of the order of r_0^6 in the expression in brackets are neglected.

If (36) is introduced, equation 2.4 (6) gives

$$\frac{d\lambda}{ds} = \frac{c_{St}^2}{R^3}\left(1-\frac{R}{r_0 c_{St}}\right) = \frac{1}{r_0 c_{St}} \frac{F+r_0^2 G}{(1+r_0^2 F+r_0^4 G)^3}$$

$$= \frac{1}{r_0 c_{St}}(F+r_0^2 G)(1-3r_0^2 F) = \frac{1}{r_0 c_{St}}[\sin S+r_0^2(4-5\sin^2 S)]. \tag{47}$$

Here terms of the order of r_0^4 have been neglected. Consequently we obtain

$$\frac{d\lambda}{ds} = \frac{1}{r_0 c_{St}}[\sin S+r_0^2(4-5\sin^2 S)]. \tag{48}$$

This equation defines the motion of the *particle*. The motion of the equivalent dipole is the average of $d\lambda/ds$. As the average of $\sin S$ is zero and of $\sin^2 S$ is $\frac{1}{2}$, we obtain

$$\frac{d\lambda}{ds} = \frac{1}{c_{St}} r_0(4-\tfrac{5}{2}) = \tfrac{3}{2} r_0 \frac{1}{c_{St}}. \tag{49}$$

The error is of the order of r_0^3.

This equation is identical with (25)

$$\frac{d\lambda}{dt} = \frac{3c\mu}{er_e^2}, \tag{50}$$

because

$$\frac{d}{dt} = v\frac{d}{ds}, \tag{51}$$

$$c_{St} = \sqrt{(ae/cmv)}, \tag{52}$$

$$\mu = \tfrac{1}{2}mv^2 r_e^3/a, \tag{53}$$

$$r_0 = r_e/c_{St}. \tag{54}$$

In a similar way the expression for T can be derived.

2.54. *On the region of validity of the perturbation method.* The perturbation method is applicable as soon as the radius of curvature ρ is small compared with the distance r from the dipole. If we put the limit to

$$\rho/r < 0{\cdot}01, \tag{55}$$

it is of interest to see how this condition restricts the use of the method in cosmic physics.

If the magnetic field H is due to a dipole with moment a at distance r, we have $H \geqslant ar^{-3}$. Further, we have, according to equation 2.2 (2′), $\rho = cmv_{\perp}/eH \leqslant cr^3p/ea$. Consequently we can be sure that the method is applicable when

$$r < 0{\cdot}1(ea/cp)^{\frac{1}{2}} = 0{\cdot}1c_{\mathrm{St}}. \tag{56}$$

The values of c_{St} for some interesting cases are given in Table 2.1.

A study of the table shows that for electrons with energy as high as about 10^5 e.v. (which probably is the upper limit of the energy of the auroral particles) our method is applicable in the earth's field almost to the moon's orbit, and in the sun's field to a distance of 10 times the radius of the earth's orbit. For protons with the same energy the region of validity is restricted to about 7 times the earth's radius, but such particles cannot be expected to be of any importance in the physics of the earth. For protons of 'thermal energy' ($\leqslant$ 1 e.v.) the limits are very large. The perturbation method is, of course, not applicable to cosmic rays with energies of 10^{10} e.v. or more (also for the reason that we have not applied relativistic mechanics).

Consequently, in almost all problems of cosmic physics—except cosmic rays—the motion of charged particles can be treated according to the perturbation method.

2.6. Cosmic-ray orbits

For cosmic rays the perturbation method is not applicable. The orbits must be determined through numerical integration of Störmer's equations. Much work has been spent on this important problem, especially by Störmer, Lemaître, and Vallarta (review, see Meixner, 1943).

It is also possible to determine the paths by a scale-model experiment. This has been done by Malmfors (1945). His results are accurate to within a few degrees, which is fully enough for all problems concerning the terrestrial magnetic field. His diagrams show from what point of the sky particles originate which reach an observer on the earth from the zenith. For latitude $\lambda = 58°$ he gives data for all incident directions.

2.7. Radiation losses

A charged particle moving in a circle emits electromagnetic radiation, which diminishes its energy. According to Larmor's formula the energy

radiated per second is
$$P = \frac{2}{3}\frac{e^2}{c^3}\left(\frac{dv}{dt}\right)^2. \tag{1}$$

For a particle moving in a circle in a magnetic field we have

$$\frac{d\mathbf{v}}{dt} = \frac{e}{mc}[\mathbf{vH}].$$

The kinetic energy of the particle is $W = \frac{1}{2}mv^2$. This gives the decay time T for the energy as

$$T = \frac{W}{P} = \frac{3}{4}\frac{c^5}{e(e/m)^3}\frac{1}{H^2}. \tag{2}$$

For one single electron this gives

$$T = 2{\cdot}55\,.\,10^8 H^{-2}, \tag{3}$$

where T is given in seconds and H in gauss. As this time is very long, the radiation losses of a single electron are usually negligible. (Only in extreme cases, such as treated by Pomeranchuk, 1940, and Tzu, 1948, it may be considerable.)

In cosmic physics we have usually to do with problems involving many particles. If, for example, several electrons move together, they radiate much more, because (1) contains the square of the charge. From the theory of magnetrons it is well known that electrons in a magnetic field have a certain tendency to 'bunch' so that a large fraction of them oscillate with the same phase. The result is an increase in radiation, so that T is very often many orders of magnitude smaller than the value given by (3). Experimental investigations by Åström (1948) on electrons drifting in crossed magnetic and electric fields have shown that they radiate a 'noise' containing frequencies distributed over a very large range. The radiated energy increases very rapidly with the density. With a magnetic field of 100 gauss and a density of about 10^8 electrons/cm.^{3}, T is less than one microsecond. Hence we must be cautious in using (2), because in cosmic physics, where we seldom have to do with one single electron, it may be in error by many orders of magnitude.

REFERENCES

General references: C. Störmer's most important papers on the motion of charged particles are: 'Sur les trajectoires des corpuscules électrisés dans l'espace sous l'action du magnétisme terrestre', *Arch. Sci. Phys., Genève*, **24**, 5, 113, 221, 317 (1907); **32**, 117, 190, 277, 415, 501 (1911); 'Ein Fundamentalproblem der Bewegung einer elektrisch geladenen Korpuskel im kosmischen Raume', *Z. f. Astrophys.*, **3**, 31, 227 (1931); **4**, 290 (1932); **6**, 333 (1933): 'On the Trajectories of Electrical Particles in the Field of a Magnetic Dipole with Applications to the Theory of Cosmic Radiation', *Avh. Vid. Akad. Mat.-nat. Kl. Oslo*, 1933,

1934; *Astrophysica Norw., Oslo,* **1,** 1, 115 (1934–5); **2,** 1, 193 (1936–7). Summaries of Störmer's theory are included in most handbooks, e.g. Chapman, C., and Bartels, J., 1940, *Geomagnetism,* **2,** 834. Compare also Hewson, E. W., 1937, 'A Survey of Facts and Theories of the Aurora', *Rev. Mod. Phys.,* **9,** 403, 1937.

Alfvén, H., 1940: 'On the Motion of a Charged Particle in a Magnetic Field', *Ark. f. mat., astr. o. fysik,* **27 A,** No. 22.

Åström, E., 1948: 'Experimental investigation on an electron gas in a magnetic field', *Trans. Royal Inst. of Technology, Stockholm,* No. 22, p. 70.

Brüche, E., 1931: 'Some New Theoretical and Experimental Results on the Aurora Polaris', *Terr. Mag.,* **36,** 41.

Malmfors, K. G., 1945: 'Determination of Orbits in the Field of a Magnetic Dipole with Application to the Theory of the Diurnal Variation of Cosmic Radiation', *Ark. f. mat., astr. o. fysik,* **32 A,** No. 8.

Meixner, J., 1943: 'Kosmische Strahlung und Magnetfeld der Erde', *Vorträge über kosmische Strahlung,* herausgegeben von W. Heisenberg, Berlin.

Pomeṙanchuk, 1940: *J. Phys. Acad. Sci.* U.S.S.R., **2,** No. 1.

Störmer, C., 1913: 'Sur un problème relatif aux mouvements des corpuscules électriques dans l'espace cosmique', *Videnskapsselskabets Skr. Mat.-nat. Klasse,* No. 14, Kristiania (Oslo).

Tzu, H. Y., 1948: 'On the Radiation emitted by a Fast Charged Particle in the Magnetic Field, *Proc. Roy. Soc.,* A, **192,** 231.

III

ELECTRIC DISCHARGES IN GASES

3.1. Introduction

ELECTRIC fields are likely to be produced, especially by induction, in stellar atmospheres and in interstellar space. Such fields accelerate charged particles, which are present wherever ionized matter exists, thus causing currents. Traditionally a current through a gas is called a discharge.

The most important cosmic phenomena which may be interpreted as electric discharges are solar prominences (§ 5.6) and aurorae (§ 6.4).

In principle it is possible to calculate the motion of electrons and ions in an electric field and in such a way predict the properties of the discharge. From laboratory studies of the discharges we know, however, that such a procedure is very difficult and dangerous. In fact, the theory must take account of all the complicated interactions between electrons, ions, molecules, and quanta. Hence it has been possible to build up the theory of electric discharges only through a very intimate contact with experiments, and many times it has been found that phenomena occur in a way which from the theoretical point of view was considered to be impossible. A striking example of this is supplied by the investigations of the ignition time of an electric spark, where a succession of theories has been disproved by experiments (see Loeb and Meek, 1941). As another example of the still precarious state of the theory we may mention the cathode mechanism of an ordinary arc. Although a phenomenon which because of its theoretical as well as overwhelming technical interest has been very much studied, no adequate theory of it exists. (In many cases thermionic as well as field emission is ruled out; see Loeb, 1939, p. 629.)

There is no reason to doubt that cosmic discharge phenomena offer problems as complicated as those occurring in laboratory discharges. A purely theoretical treatment of them is certainly very precarious, and still, in many cases, this is the only way of attack.

One important approach to the study of electric currents in cosmic physics is due to Chapman and Cowling. From the mathematical theory of gases they have deduced the conductivity of ionized gases (see Chapman and Cowling, 1939). In view of what is said above an empirical check is highly desirable.

Another, and quite different, line of approach was tried by Birkeland

(1901). His famous terrella experiment was an attempt to solve cosmic discharge problems by model experiments. The investigation has had a very inspiring effect on cosmic physics, but, as it was made before the theory of gaseous discharges, no one really knows what happened in his vessel. Probably the interpretation of his experiment was inadequate (see § 6.1).

Since the time of Birkeland's experiments the theory of electric discharges has been developed very much, but until quite recently no attempt had been made to transfer the results to the realm of cosmic physics. Certainly many parts of the theory are still unsatisfactory, but, for example, our knowledge of the properties of a plasma is so good that valuable results may be expected if we apply it to cosmic problems. What is urgently needed at present is not a refined mathematical treatment but a rough analysis of the basic phenomena. The purpose of this chapter is to draw attention to some aspects of discharge theory which may be important in cosmic physics.

3.11. *Survey of electric discharges.* Electric discharges are usually divided into two groups: *non-sustained discharges,* which are dependent upon an 'external' ionizer producing at least an essential part of the ions and electrons which carry the current, and *self-sustained discharges,* where the ionization is mainly produced by the discharge itself. *Ceteris paribus* the second is characterized by higher current densities than the first. This is due to the fact that in the laboratory we have at our disposal only very weak ionizers. In cosmic physics, where the 'external ionizer' may be a high temperature which ionizes the matter more or less completely, non-sustained discharges may carry very large currents.

The domain of the self-sustained discharges is very extensive, including Townsend discharges, glow discharges, and arcs. Moreover, there are several special forms such as the spark, which is essentially a short-lived arc. In most of the discharges we can discern three different regions:

1. The cathode region, where the electrons (which carry the main part of the current) are produced by emission from the cathode and by ionization of the gas.
2. The anode region (which is rather unimportant), associated with the passing of the current between the discharge and the anode.
3. The 'plasma' which extends from the region of the cathode mechanism to that of the anode mechanism. The properties of the plasma can be regarded as characteristic for a gaseous conductor in the absence of disturbances from electrodes.

The distinction between the different types of discharges lies mainly in the cathode mechanism. In the Townsend and the glow discharge the emission takes place from a cold cathode; in the arc the cathode is hot enough to give thermionic emission (or it emits abundantly for some other reason).

The properties of the plasma are not immediately connected with the cathode mechanism, so in principle the plasma could have the same properties for all types of self-sustained, and even for non-sustained, discharges. The state of the plasma depends upon the current density, and this is usually increasing when we go from non-sustained to Townsend and further to glow and to arc discharges. Although in principle the same phenomena occur in all plasmas, the properties of an arc plasma are, because of the high current density, different from that of a glow discharge, and still more different from that of a non-sustained discharge.

In cosmic physics the cold cathode mechanisms are of little interest. If we can speak of electrodes these usually consist of ionized gaseous layers of higher density than the discharge space. Such layers can give off electrons abundantly, so that the cathode mechanism is most similar to that of an arc discharge.

3.12. *Similarity transformations.* In the theory of gaseous discharges certain 'similarity laws' have proved very valuable (see Cobine, 1941, p. 209, or Engel and Steenbeck, **2**, 1934, p. 95). When changing the linear scale by a factor η the most characteristic features of the phenomena remain unchanged if at the same time we change other quantities according to Table 3.1.

TABLE 3.1

Similarity transformation applicable to gaseous discharges

Length, time, inductance, capacity	vary as η^{+1}
Particle energy, velocity, potential, current, resistance	,, as η^{0}
Electric and magnetic field, conductivity, gaseous density	,, as η^{-1}
Current density, space charge density	,, as η^{-2}

Proportionality between length and time is required by Maxwell's equations. The most characteristic features of a discharge depend upon the interactions between atoms, electrons, and quanta. As these interactions depend in a very complicated way upon the energies involved, we must leave all energies, and hence the electrostatic potential (which determines the kinetic energy of a charged particle) unchanged. If we change the linear dimensions l by a factor η, the electric field E must be changed by η^{-1} in order to leave the potential $V \sim lE$ unchanged.

Because of Maxwell's equations we must change D, H, and B in the same way as E. The current density i which is equivalent to the displacement current $\partial D/\partial t$ must be changed by the factor η^{-2}, which means that the total current $I = il^2$ remains unchanged. The conductivity σ ($= i/E$) changes as η^{-1}, the inductance L, which equals $V/(dI/dt)$, and the capacity C ($\sim l$) change as η. Further, as the mean free path, which is of fundamental importance in gaseous discharges, varies as the linear dimension, the density ρ of the gas, which is inversely proportional to the mean free path, must be changed as η^{-1}.

In the theory of gaseous discharges the above transformation has proved to be very useful in making a general survey, but it must be used with some care, because it refers to the most fundamental phenomena only, and many secondary phenomena, which in special cases become important, do not obey the transformation. For example, the number of charged particles per unit volume is proportional to i and hence varies as η^{-2}, whereas the number of molecules is proportional to ρ and hence varies as η^{-1}. Hence the degree of ionization is not invariant, as we should like it to be, but varies as η^{-1}. Further, as the force f ($= iH/c$), which acts upon unit volume traversed by a current i in the presence of a magnetic field H, is proportional to η^{-3}, but the density is proportional to η^{-1}, the acceleration becomes proportional to η^{-2}, and not, as it ought to be because of its dimension lt^{-2}, to η^{-1}. One of the consequences of this is that magneto-hydrodynamic waves (see Chapter IV) do not obey the transformation.

It must be observed that the transformation does not affect atomic quantities. For example, atomic dimensions, wave-length of emitted light, and lifetime of metastable states will remain unchanged.

If we want to apply the results obtained in a laboratory apparatus with the linear extension of 10 cm. to cosmic phenomena, we have to increase the scale by a factor of 10^8–10^9 with regard to the conditions around the earth, a factor of 10^{10} for the sun, 10^{12}–10^{13} for the planetary system, and 10^{21}–10^{22} for the galaxy. Perhaps it is of more interest to go the other way, i.e. to transform the cosmic phenomena down to laboratory scale, because this gives us some hint concerning the general type of the phenomena. It shows what quantities are the most important ones, and indicates to what extent it is possible to make scale-model experiments illustrating cosmic phenomena.

Table 3.2 shows how the similarity transformation may be applied to some important domains of cosmic physics.

The table shows some features of interest. The first is that most

TABLE 3.2

Problem	*Linear dimension*	*Density particles/cm.*^{3}	*Magnetic field gauss*	*Time*
Aurora and magnetic storms . . . Reduced: $\eta = 3.10^8$.	3.10^9 10	10^8 ?–10^{12} 3.10^{11} ?–3.10^{20}	0·5–0·01 $1·5.10^8$–3.10^6	Initial phase of storm = 3h. = 10^4 sec. → 30 μsec.
Solar corona . . . Reduced: $\eta = 10^{10}$–10^{11}	10^{11}–10^{12} 10	10^8–10^6 10^{18}–10^{17}	20–0·02 2.10^{11}–2.10^9	Life of coronal arc = 10^3 sec. → 10^{-7} sec. Solar cycle = 11 years = 3.10^8 sec. → 0·03 sec.
Chromosphere . . Reduced: $\eta = 10^8$.	10^9 10	10^{11}–10^{14} 10^{19}–10^{22}	20 2.10^9	Solar flare 1,000 sec. → 10 μsec. Prominence 10^5 sec. → 1,000 μsec.
Planetary system . . Reduced: $\eta = 10^{12}$–10^{13}	10^{13}–10^{14} 10	10^3 ? 10^{15}–10^{16} ?	10^{-5}–10^{-8} 10^7–10^5	1 year → 3–30 μsec.
Galaxy . . . . Reduced: $\eta = 3.10^{21}$.	3.10^{22} 10	1 3.10^{21}	10^{-12} ? 3.10^9 ?	Age of universe = 10^{10} years = 3.10^{17} sec. → 100 μsec.

At normal temperature a density of $3·6.10^{16}$ particles/cm.^{3} corresponds to a gas pressure of 1 mm. Hg.

The density in the planetary system is a guess founded on the reasonable assumption that the density must be intermediate between the coronal and the interstellar values. The same value is used for 'aurora and magnetic storms', representing the density at some distance from the earth. The latter value in the same square refers to the upper atmosphere (*E*-layer).

pressures are to be considered as rather high. In the case of the planetary system a reduced value of some tenth of a millimetre is obtained, but this is, of course, rather uncertain. At a display of aurora and magnetic storm, the density around the earth corresponds to about 10^{-5} mm. Hg. All other equivalent pressures are well above 1 mm. Hg ($3·6.10^{16}$ particles cm.$^{-3}$). Consequently with the above exception there is no analogy to high-vacuum phenomena in cosmic physics. When considering electrical phenomena the interstellar space of our galaxy should not be compared with a 'vacuum' but with a highly ionized gas at a pressure of 100 atmospheres.

Still more striking than the high densities are the very high magnetic fields in the cosmos. In fact, they are so strong that at present our laboratory resources do not suffice to produce fields strong enough for model experiments.

The powerful magnetic fields have two important consequences. The first is that the motion of charged particles is usually of a different type from what we are familiar with in the laboratory. The radius of curvature is very small and the particles move in the direction of the magnetic

field or 'drift' perpendicular to it. This type of motion has been studied in Chapter II.

The other consequence is that strong electric fields are easily produced by any motion across the magnetic field (see § 1.3). To give an example, in a magnetic field of 10^6 gauss, a velocity of 3.10^5 cm./sec. causes an electric field of $E = 10$ e.s.u. $= 3{,}000$ volt/cm., and in a field of 10^{10} gauss the same velocity gives 30.10^6 volt/cm. Thus the electric fields in the cosmos also, when reduced to laboratory scale, are often very strong.

Finally, the time-scale transformation in Table 3.2 is of interest. Solar flares, coronal arcs, and also the initial phase of a magnetic storm should be regarded as very short-lived phenomena. In fact their equivalent duration (a few μsecs.) is of the order of the ignition time of an electric discharge.

3.13. *Properties of a plasma.* From what has been said in § 3.11, it is evident that from the point of view of cosmic physics a survey of the properties of a plasma is of special interest. Because of the importance of the cosmic magnetic fields we must also pay attention to the influence of magnetic fields upon a discharge, a phenomenon which has not been studied very much in the laboratory.

A plasma consists of neutral molecules (monatomic or polyatomic), electrons, positive (and in many cases also negative) ions, and also quanta, emitted from the excited atoms. The presence of an electric field is essential. In most laboratory discharges the degree of ionization is very small. In cosmic physics the ionization may be more or less complete in many cases.

The electrons, ions, and molecules collide mutually. In a typical plasma only a very small fraction of the electrons have velocities so large that they can ionize or excite the molecules. Hence most collisions between electrons and molecules are elastic. Due to the big difference in mass between electrons and other particles, the exchange of energy is small at such a collision. In fact an electron transmits only a fraction of the order of m_e/M (m_e = electronic, M = molecular mass) of its kinetic energy when colliding with a heavy particle. Hence if the mean energy of the electrons is different from that of the molecules, several thousand collisions are required before the energies are equalized (m_e/M being $\leqslant 1/1840$). On the other hand, the ions and molecules have masses of the same order, so that at collisions the energy exchange is of the same order as the total kinetic energy. A difference in mean energy is rapidly smoothed out.

In a plasma the velocity distribution of the molecules is, at least to a

first approximation, Maxwellian, as in an ordinary gas. We call its temperature T_M. The ions and electrons are affected by the electric field, which gives them a systematic velocity parallel or antiparallel to the field. In a typical plasma this velocity is small in comparison to the random velocity, which even for the ions and electrons to a first approximation is Maxwellian. Thus we can speak of an electronic gas having a certain 'electronic temperature' T_e which is defined through the condition that $3/2kT_e$ (k = Boltzmann's constant) shall equal the average energy due to the random velocity of the electrons. In the same way there is an ionic gas having the 'ionic temperature' T_i.

The systematic motion in the electric field causes a heating of the electronic gas as well as of the ionic gas. Due to the small energy exchange between the electrons and the other constituents, the electronic gas may reach a temperature which is one or two (or even three) powers of 10 above that of the molecular gas. On the other hand, the thermal contact between the ionic and the molecular gas is good enough to ensure that no big difference between the ionic and molecular temperature is established.

The average energy of the electrons is usually much lower than the ionization or excitation energies of the molecules. Only that small part of the electrons which, due to the Maxwellian distribution, have energies several times the average energy are able to ionize or excite.

The behaviour of a plasma is very complicated because so many different reactions are possible between electrons, more or less excited or ionized atoms or molecules, and quanta. In principle it is possible to treat the phenomena by exact statistical methods. In practice, however, most theories must be approximate because of the complexity of the problems. For several purposes we reach sufficient accuracy without using the more elaborate methods of statistical mechanics.

3.2. Mobility and conductivity

When an electric field is applied to an ionized gas, all charged particles are accelerated. Because of the friction with the rest of the gas, their average velocities in the direction of the electric force soon reach stationary values. Suppose that the volume density of one kind of charged particles with charge e_k and mass m_k is n_k. An electric field E gives to these particles an average 'drift' velocity u_k:

$$u_k = b_k E. \tag{1}$$

For weak fields b_k is a constant called the *mobility*. When the mobilities

of all the constituents of a gas are known, its *conductivity* σ can easily be computed. In fact, the current density, i, produced by the field is given by

$$i = \sum_k n_k e_k u_k = E \sum_k n_k e_k b_k, \tag{2}$$

where the summation includes all kinds of charged particles. As the conductivity is defined by

$$\sigma = i/E, \tag{3}$$

we have

$$\sigma = \sum_k n_k e_k b_k. \tag{4}$$

For negative particles e as well as b is negative.

Using exact statistical methods Chapman and Cowling have calculated the conductivity of ionized gases (see Chapman and Cowling, 1939). Their calculations are mathematically difficult, so for a survey it is preferable to use the simpler 'free path method' which more easily demonstrates the physical process. As the authors just cited have shown, this gives, in general, formulae which are sufficiently accurate in view of the fact that the final results anyhow depend on uncertain parameters, such as atomic collisional cross-sections.

In the free path method it is assumed that the molecules (including ions and electrons) of a gas make instantaneous collisions with each other, but move freely between the collisions. In a rigorous treatment according to this method the statistical distribution of molecular velocities, free paths, etc., is taken into consideration, but in a less accurate variant of the method only the average values of the quantities are used. The results differ in general by less than a factor of 2, an accuracy which is enough for most applications.

Let v_k be the mean temperature velocity, λ_k the mean free path, and τ_k the mean interval between two collisions which a particle of kind k makes with other particles:

$$\tau_k = \lambda_k/v_k. \tag{5}$$

During the time τ_k the particle falls freely in the electric field a distance

$$D_k = \frac{e_k E}{2m_k}\tau_k^2. \tag{6}$$

Hence its average motion in the direction of the electric field is

$$u_k = \frac{D_k}{\tau_k} = \frac{e_k E}{2m_k}\tau_k = \frac{e_k \lambda_k}{2m_k v_k} E. \tag{7}$$

This gives the mobility

$$b_k = u_k/E = \gamma e_k \lambda_k/m_k v_k \tag{8}$$

with $\gamma = \frac{1}{2}$. A more rigorous treatment taking account of the statistical

distribution of velocities and free paths gives the same formula but with $\gamma = 1$. Different authors give values of γ between these (see Cobine, 1941, p. 33). For electrons we may put $\gamma = 0{\cdot}85$.

For low values of E, v_k and hence b_k are independent of E, but when E is greater than a certain value, they vary as we shall now see.

When the particle drifts in the electric field, its kinetic energy W_k,

$$W_k = \tfrac{1}{2}m_k v_k^2, \tag{9}$$

increases at the rate
$$dW_k/dt = e_k E u_k. \tag{10}$$

Let κ_k be the fraction of its kinetic energy which it loses on an average when colliding with a molecule. A stationary state is reached when the increase in energy due to the drift in the electric field equals the energy losses at the $1/\tau$ collisions per second:

$$\frac{dW_k}{dt} = \frac{\kappa_k W_k}{\tau_k}. \tag{11}$$

Combining (10) and (11), and introducing (5), (8), and (9), we obtain

$$\tfrac{1}{2}\gamma e_k^2 \lambda_k^2 E^2 = \kappa_k W_k^2, \tag{12}$$

or
$$W_k = \sqrt{(\gamma/2\kappa_k)}e_k \lambda_k E. \tag{13}$$

The value of the collision loss ratio κ_k depends upon the character of the collision. If it is inelastic κ_k may be as high as 1, the total kinetic energy of the colliding particle being transformed into excitation or ionization energy. Usually most collisions are elastic, and in this case κ_k can be computed from a formula given by Cravath (1930):

$$\kappa_k = \gamma_1 \frac{m_k}{M}\left(1 - \frac{W_M}{W_k}\right), \tag{14}$$

where M is the molecular mass, W_M the average kinetic energy of a molecule, and γ_1 is a constant of order unity:

$$\gamma_1 = \frac{8}{3}\left(\frac{M}{m_k+M}\right)^2. \tag{15}$$

Introducing (14) into (12) we find

$$W_k = \tfrac{1}{2}W_M + \sqrt{\{\tfrac{1}{4}W_M^2 + \gamma_2 M(e_k \lambda_k E)^2/m_k\}}, \tag{16}$$

where
$$\gamma_2 = \tfrac{1}{2}\gamma/\gamma_1. \tag{17}$$

γ_2 is a numerical constant. If the charged particles are ions ($k \to i$) so that $M/m_i \sim 1$, the average energy, W_i, of these is about the same as the average molecular energy W_M unless E is very large. If the charged particles are electrons ($k \to e$) the ratio $M/m_e \geqslant 1{,}840$, so that even relatively weak fields make W_e considerably larger than W_M. This is in

agreement with what was stated in § 3.13: the ionic temperature T_i is usually approximately equal to the gas temperature, whereas the electronic temperature T_e easily becomes much higher. As we have

$$W_M = \tfrac{3}{2}kT_M \tag{18}$$

and, if the charged particles are electrons,

$$W_e = \tfrac{3}{2}kT_e, \tag{19}$$

we can write for (16):

$$T_e = \tfrac{1}{2}T_M+\sqrt{\{\tfrac{1}{4}T_M^2+\gamma_2 M(\epsilon\lambda_e E)^2/m_e\}}, \tag{20}$$

where ϵ is the conversion factor between electron volts and temperature

$$\epsilon = 2e/3k = 2{\cdot}32.10^6 \text{ e.s.u.} = 7{,}700 \text{ degrees/e.v.} \tag{21}$$

The electronic temperature becomes considerably different from the gas temperature when $\lambda_e E$ increases beyond the value given by

$$\lambda_e E = \frac{1}{\epsilon}\sqrt{\left(\frac{m_e}{M}\right)}T_M. \tag{22}$$

If $M = 2{,}000m_e$, this occurs for $T_M = 300°$ when the product of the mean free path and the electric field, $\lambda_e E$, equals 0·001 volt, and for $T_M = 6{,}000°$ when $\lambda_e E = 0{\cdot}02$ volt. In laboratory discharges, electron temperatures of 20,000–50,000° are frequently measured.

As long as $T_e \approx T_M$, the drift is proportional to the electric field E. According to (8) it is given by:

$$u_e = \gamma\frac{e}{m_e}\frac{\lambda_e}{v_e}E, \tag{23}$$

where $\gamma \approx 1$, and m_e means the electronic mass; v_e is given by

$$\tfrac{1}{2}m_e v_e^2 = \tfrac{3}{2}kT_e \approx \tfrac{3}{2}kT_M. \tag{24}$$

If instead $T_e \gg T_M$, we have approximately (according to (20))

$$T_e = \epsilon\sqrt{(\gamma_2 M/m_e)}.\lambda_e E. \tag{25}$$

From (9), (19) and (21) we find the heat motion velocity

$$v_e = (e/m)^{\frac{1}{2}}(4\gamma_2 M/m_e)^{\frac{1}{4}}(\lambda_e E)^{\frac{1}{2}}. \tag{26}$$

Introducing (26) into (8) we find that the drift velocity

$$u_e = (e/m_e)^{\frac{1}{2}}.(\tfrac{1}{2}\gamma^3\gamma_1)^{\frac{1}{4}}(m_e/M)^{\frac{1}{4}}(\lambda_e E)^{\frac{1}{2}} \approx (m_e/M)^{\frac{1}{2}}v_e. \tag{27}$$

u_e is proportional to the square root of the electric field.

When the drift velocities of all charged particles are known, the current density i is found from (2).

As we have seen above, for values of the electric field E so small that the charged particles temperatures approximately equal the gas

temperature, u_i and u_e are proportional to E. In this case the conductivity is given by

$$\sigma = \gamma e^2\left(\frac{n_i \lambda_i}{m_i v_i} + \frac{n_e \lambda_e}{m_e v_e}\right). \tag{28}$$

R. 54

As $\lambda_e > \lambda_i$,? $m_e v_e \ll m_i v_i$, and usually $n_i \approx n_e$ (see § 1.4), the electronic conductivity is usually much higher than the ionic conductivity, so that the first term can be neglected:

$$\sigma = \gamma \frac{e^2}{m_e} \frac{n_e \lambda_e}{v_e} = \gamma \frac{e^2}{m_e} n_e \tau_e \tag{29}$$

($\tau_e = \lambda_e/v_e =$ time between two collisions). If the electron temperature is higher than the gas temperature, i according to (27) is proportional to $E^{\frac{1}{2}}$. If we still want to use (3) we must put σ proportional to $E^{-\frac{1}{2}}$. Thus the conductivity is independent of E for small values of E, but as soon as E increases beyond the value given by (22), the conductivity begins to decrease.

3.21. *Influence of a magnetic field.* A magnetic field parallel to the electric field does not directly affect the mobility and conductivity. As it impedes the sideways diffusion of charged particles, it often changes the degree of ionization and thus indirectly affects the conductivity in an electric discharge (see Engel and Steenbeck, **2,** 1934, p. 112). On the other hand, when the magnetic field H is perpendicular to the electric field E, new effects are produced which shall be discussed here. We suppose E and H to be homogeneous.

Consider at first a completely ionized gas which consists of electrons and positive ions and is on an average electrostatically neutral. The drift of a charged particle in combined electric and magnetic fields may be found from equation 2.2 (23). When a *stationary state* is reached f^i becomes zero because of 2.2 (24). As $f = eE$, we have

$$\mathbf{u}_\perp = -\frac{c}{H^2}[\mathbf{HE}] \tag{30}$$

independent of the charge and mass of the particle. This means that all the charged particles drift with the same velocity which is perpendicular to E and H. The current is zero because the gas is electrostatically neutral. Consequently, *as soon as a stationary state* is reached, an electric field produces no current in the presence of a transverse magnetic field.

This holds also when the gas is incompletely ionized so that it contains neutral molecules. In fact, even in this case, the charged particles will tend to drift according to (30) and they will bring the rest of the gas into

the same state of motion. The only difference is that the time required to reach the stationary state is longer.

As in the presence of a transverse magnetic field an electric field produces no current in the stationary state, it would be reasonable to say that the conductivity is zero. Conventionally, however, the 'cross-conductivity' $\sigma_\perp$ is defined by

$$\mathbf{i}_E = \sigma_\perp\{\mathbf{E}+(1/c)[\mathbf{vH}]\}, \tag{31}$$

where i_E is the current in the electric field direction and v represents the average state of motion of the gas. In the stationary state we have $v = u_\perp$, so that the bracket cancels and we obtain $i_E = 0$. This means that $\sigma_\perp$ refers to a transient current that starts at the moment when the electric field is switched on and later decays exponentially until the stationary state is reached. This current has one component parallel to E and this is defined by (31), but it has also another component i_H which is perpendicular to E as well as to H. This component is called the Hall current. The 'Hall conductivity' σ_H is defined by

$$\mathbf{i}_H = \frac{\sigma_H}{H}\left[\mathbf{H}\left\{\mathbf{E}+\frac{1}{c}[\mathbf{vH}]\right\}\right], \tag{32}$$

or

$$\frac{i_H}{i_E} = \frac{\sigma_H}{\sigma_\perp}. \tag{33}$$

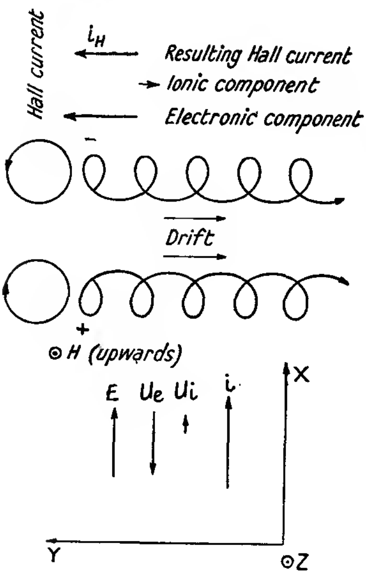

FIG. 3.1. In an electric field E positive ions move with velocity u_i and electrons with velocity u_e, producing a current i. In the presence of a magnetic field H (upwards through paper) the particles also drift perpendicular to E and H, thus causing a Hall current i_H.

The transient current is due to the differential motion of charged particles before the whole gas has been accelerated. It may be computed by introducing into equation 2.2 (23) the friction produced by the relative motion. Still adopting an approximate method we confine ourselves to the electronic part of the current which is the most important part. Suppose that at a certain instant the average drift of the electrons is $u_\perp$ relative to the average motion of the rest of the gas. At the collisions which an electron makes with the rest of the gas its average momentum before the collision is $mu_\perp$ and after the collision zero. The frictional force equals the change in momentum multiplied

by the number of collisions per second. Hence an electron is subject to the force

$$\mathbf{f} = e\mathbf{E}' - m\mathbf{u}_\perp/\tau, \tag{34}$$

where $\mathbf{E}' = \mathbf{E}+c^{-1}[\mathbf{vH}]$ and is the field measured in a system sharing the average gas motion. Introducing (34) into equation 2.2 (23) and still neglecting f^i because the change in the average drift is slow, we obtain

$$\mathbf{u}_\perp = -\frac{c}{eH^2}\left[\mathbf{H}\left\{e\mathbf{E}' - \frac{m\mathbf{u}_\perp}{\tau}\right\}\right] \tag{35}$$

and

$$\mathbf{i} = ne\mathbf{u}_\perp. \tag{36}$$

Suppose that H is parallel to the z-axis and E to the x-axis of an orthogonal reference system. Let u_E and i_E be the drift and current parallel to E, and u_H and i_H the corresponding y-components. Then we have from (35)

$$u_E = -\frac{c}{eH}\frac{mu_H}{\tau}, \tag{37}$$

$$u_H = -\frac{c}{eH}\left(eE' - \frac{mu_E}{\tau}\right). \tag{38}$$

We introduce the gyrofrequency

$$\omega_e = \frac{|e|H}{m_e c}. \tag{39}$$

We obtain from (37), (38), (31), and (32)

$$\sigma_\perp = \frac{n_e e^2\tau_e}{m_e(1+\omega_e^2\tau_e^2)} \tag{40}$$

and

$$\sigma_H = \omega_e\tau_e\sigma_\perp. \tag{41}$$

For an exact derivation of these results see Chapman and Cowling (1939). Taking also the ionic component into consideration Cowling (1945) writes:

$$\sigma_\perp + j\sigma_H = \frac{n_e e^2\tau_e^M}{m_e(1-j\omega_e\tau_e^M)} + \frac{n_i e^2\tau_i^M}{m_i(1+j\omega_i\tau_i^M)}, \tag{42}$$

where $j = \sqrt{(-1)}$, ω_i is the ionic gyrofrequency, and τ_e^M and τ_i^M the collisional times for electrons and ions with molecules.

From equations (39), 3.2 (5), and 2.2 (2), and because statistically $v_\perp = \sqrt{\tfrac{2}{3}}\,v$, we have

$$\omega\tau = \frac{eH}{mc}\frac{\lambda}{v} = \frac{v_\perp}{v}\frac{\lambda}{\rho} = \sqrt{\frac{2}{3}}\frac{\lambda}{\rho}. \tag{43}$$

Hence $\omega\tau$ is a measure of the ratio between the mean free path and the radius of curvature.

When the drift perpendicular to the electric field is braked, the

conductivity increases. When no drift is possible the cross-conductivity equals the parallel conductivity. This can be shown in the following way.

Suppose that in an electric field no Hall current i_H is allowed to flow. This may be effected, for example, by inserting insulating planes parallel to the electric and magnetic fields. Then i_H is compensated by a conduction current i', which is due to a secondary field produced by the impeded Hall current. This secondary field also produces a Hall current i'_H. We write $i_H = \omega_e \tau_e i$, or in vector notation:

$$\mathbf{i}_H = \omega_e \tau_e (1/H)[\mathbf{Hi}],$$

and have

$$\mathbf{i}' + \mathbf{i}_H = 0,$$

$$\mathbf{i}'_H = \omega_e \tau_e (1/H)[\mathbf{Hi}'].$$

The result is a current i'' in the direction of the primary field E:

$$\mathbf{i}'' = \mathbf{i} + \mathbf{i}'_H = i + \omega_e^2 \tau_e^2 \mathbf{i} = (1 + \omega_e^2 \tau_e^2) \sigma_\perp \mathbf{E},$$

or simply

$$i'' = \sigma_\parallel E.$$

If the Hall current is prohibited the conductivity is independent of the magnetic field.

The cross-conductivity $\sigma_\perp$ never enters at a stationary state. If the medium cannot move freely, it is replaced by the parallel-conductivity $\sigma_\parallel$. If the medium can move freely, the cross-conductivity certainly gives a current but only for a time, after which the matter is accelerated to such a velocity that its polarization compensates the electric field and no current flows perpendicular to the magnetic field. In problems of this kind magneto-hydrodynamic waves may be produced (cf. Chap. IV).

3.22. *Equivalent circuit.* The general analogy between a mechanical system and an electric circuit makes it possible to survey the behaviour of a gas by drawing its equivalent electric circuit. We start with the simple case of a homogeneous conducting medium with density ρ and conductivity σ, which is freely movable and acted upon by crossed electric and magnetic fields E and H. The equivalent circuit of a unit cube of the medium consists of a resistance R in series with a condenser C (Fig. 3.2 *a*). The applied voltage is V, the condenser voltage V_c, and the current I. It is easily seen that we have the following correspondence

$$\begin{aligned} I &\to i \\ V &\to E \\ V_c &\to vH/c \\ R &\to 1/\sigma \\ C &\to \rho c^2/H^2. \end{aligned} \tag{44}$$

The charging of the condenser corresponds to the acceleration of the medium by the current density i. The magnitude of the condenser is chosen so as to make its electrostatic energy $\frac{1}{2}CV_c^2$ equal to the kinetic energy $\frac{1}{2}\rho v^2$ of the medium.

When the medium is a gas consisting of charged particles the influence

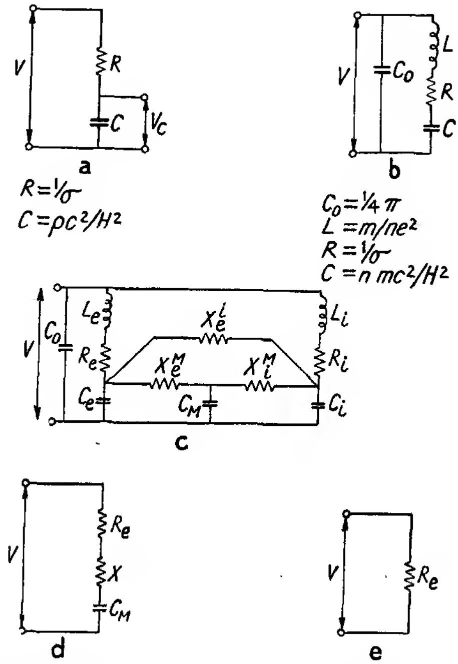

FIG. 3.2. Equivalent circuits of conductors in a magnetic field. *a*. Conducting liquid. R represents the specific resistance; the energy of condenser C represents the kinetic energy. *b*. Charged particle gas. L represents the inertia of the charged particles. C_0 represents the capacity in vacuum. *c*. Complete circuit of ionized gas. L_e, R_e, C_e represent the electronic gas, L_i, R_i, C_i the ionic gas, and C_M the molecular gas. The resistances X_e^M, X_e^i, X_i^M represent the friction between the constituents. *d*. Simplified circuit of ionized gas in crossed fields. *e*. Circuit of gas in parallel fields.

of the inertia of the particles (mass $= m$, density $= n$) is represented by a series inductance L,

$$L \to \frac{m}{ne^2}. \tag{45}$$

The resonance frequency of this inductance and the capacity C,

$$C \to \frac{\rho c^2}{H^2} = \frac{nmc^2}{H^2}, \tag{46}$$

is the gyrofrequency $\omega = (LC)^{-\frac{1}{2}} = eH/mc$. In order to account even

for the displacement current $i = 1/4\pi\, dE/dt$ we must introduce a condenser $C_0 = 1/4\pi$. The circuit is seen in Fig. 3.2 *b*.

Consider an ionized gas, consisting of molecules, electrons, and ions. The equivalent circuit is composed of the circuits for each of the constituents (see Fig. 3.2 *c*). The electronic gas is represented by L_e, C_e, and R_e, the ionic gas by L_i, R_i, and C_i. As the conductivity of the molecular gas is zero, this constituent is represented by the condenser C_M only, the energy of which corresponds to the kinetic energy of the molecules, when drifting perpendicular to the field. If the gas is electrostatically neutral ($n_e = n_i$) we have, according to (46):

$$C_i = \frac{m_i}{m_e} C_e, \tag{47}$$

so that C_i is much bigger than C_e. If the degree of ionization (= relative number of ionized atoms) is low, C_M is bigger than C_i. Except when the ionization is almost complete, we have

$$C_M \gg C_e. \tag{48}$$

The friction between the gases is represented by the resistances X_e^i, X_e^M, and X_i^M. In order to find the value of X_e^M, let us consider the circuit C_e—X_e^M—C_M alone. The time constant τ of this circuit is given by $\tau = C_e C_M X_e^M/(C_e + C_M)$ or (because $C_e \ll C_M$) approximately $\tau = C_e X_e^M$. If the voltage of C_e differs from that of C_M, this corresponds to a difference in velocity between the electronic gas and the molecular gas. The drift velocities become equalized when every electron has collided once with a molecule. Denoting the average time between two collisions which an electron makes with a molecule by τ_e^M, we have an equivalence between the time constant τ and τ_e^M. Consequently we may put

$$\tau_e^M = C_e X_e^M, \tag{49}$$

which gives [compare equations (46) and 3.21 (39)]

$$X_e^M = \frac{\tau_e^M H^2}{n_e m_e c^2} = \frac{m_e \omega_e^2 \tau_e^M}{e^2 n_e}. \tag{50}$$

Analogous expressions are obtained for X_e^i and X_i^M. (Note that $n_i \tau_M^i = n_M \tau_i^M$ and that a molecule when colliding with an electron transmits on an average only m_e/m_M of its momentum.) Except in extreme cases X_i^M is much smaller than X_i^e or X_e^M.

In the case of an electrostatically neutral slightly ionized gas

$$(n_e = n_i \ll n_M),$$

which is one of the most important cases, the dominating elements of

the circuit are C_M, R_e, and X_e^M. To a first approximation we may neglect all other elements, so that the equivalent circuit becomes as shown in Fig. 3.1 *d*. The field gives energy to the electrons (through R_e) which in their turn transmit it to the atoms (through X_e^M). If for the sake of simplicity we put $\gamma = 1$ in (29) we obtain:

$$R_e = \frac{m_e}{e^2 n_e \tau_e}. \tag{51}$$

As the electrical conductivity should be the inverse value of the resistance in series with the condenser, the circuit of Fig. 3.1 *d* gives for the cross-conductivity $\sigma_\perp$ in the magnetic field

$$\sigma_\perp = \frac{1}{R_e + X} \tag{52}$$

(where X_e^M is abbreviated to X), or because of (50) and (51),

$$\sigma_\perp = \left[\frac{m_e}{e^2 n_e \tau_e} + \frac{m_e \omega_e^2 \tau_e}{e^2 n_e}\right]^{-1} = \frac{e^2 n_e \tau_e}{m_e(1+\omega_e^2 \tau_e^2)} \tag{53}$$

where τ_e is substituted for τ_e^M because we have assumed the degree of ionization to be small. This is the same result as obtained in equation 3.21 (40).

For densities so high that the mean free path λ is much smaller than the radius of curvature ρ, we have $R_e \gg X$. Hence the contact between the electronic and molecular gas is good so that their states of motion are approximately the same. The resistance is due to the difficulty of transmitting energy from the field to the electrons. For small densities ($\lambda \gg \rho$), we have $X \gg R_e$. The energy is easily transmitted from the field to the electronic gas, which possesses almost the full drift velocity ($V_c \approx V$), whereas the transmission of energy from the electronic gas to the molecular gas is a slow process.

3.23. *Mean free path. Numerical values of conductivity.* The quantities entering into the conductivity formulae 3.2 (29), 3.21 (40), and 3.21 (41) are atomic constants (e, m_e), the density n_e, the gyrofrequency ω_e, and the collisional time τ_e. Only the last one is difficult to calculate. According to 3.2 (5) we have

$$\tau_e = \lambda_e / v_e, \tag{54}$$

where the velocity v_e is given by

$$v_e = (3kT_e/m_e)^{\frac{1}{2}}. \tag{55}$$

Here k is Boltzmann's constant, m_e the electronic mass, and T_e the electronic temperature. [In an accurate analysis the difference between the root mean square velocity $(3kT_e/m_e)^{\frac{1}{2}}$, and average velocity $(8kT_e/\pi m_e)^{\frac{1}{2}}$ ought to be taken into consideration.] The problem is to determine λ_e.

The mean free path is a simply definable quantity when the molecules are supposed to collide as elastic balls. For collisions between equal balls with the same average velocity and the same cross-section S_M, we have

$$\lambda_M = \frac{1}{nS_M}, \tag{56}$$

where n is the number of balls per unit volume. If electrons are supposed to be balls much smaller than the molecules and to move much quicker, their mean free path is given by $\lambda_e = 4\sqrt{2}\lambda_M$.

In reality the phenomena are much more complex, because the interaction between the particles is not confined to certain instants of collision. If we still want to use the same terms, and put

$$\lambda_e = \frac{1}{nS}, \tag{57}$$

S must be introduced as a complicated function of the electronic velocity and hence the temperature. Diagrams and tables of S are found in most text-books (e.g. Engel–Steenbeck **1**, p. 168; Cobine, p. 29). Usually the function $3{\cdot}6.10^{16}S$ is given, which is the sum of the cross-sections of all molecules contained in a cubic centimetre of a gas at 1 mm. Hg and 0°. The inverse value gives the mean free path at this pressure and temperature. For electrons below some hundred e.v. the cross-section, S, of most gases is of the order of magnitude of 10^{-15} cm.^{2} Introducing (57) into 3.2 (29) we find ($\gamma \approx 1$)

$$\sigma_{\parallel} = \frac{e^2}{m_e}\frac{1}{S}\frac{1}{v_e}\frac{n_e}{n}. \tag{58}$$

For a completely ionized gas the cross-section S refers to the collisions between electrons and ions and may be roughly estimated in the following way. When an electron approaches an ion, the main force acting upon it derives from the Coulomb field. At the distance x from a Z-fold ionized atom, the electrostatic energy of an electron is

$$w = \frac{Ze^2}{x}. \tag{59}$$

If this energy is a considerable fraction—say the fraction γ'—of the kinetic energy W_e ($= \frac{3}{2}kT_e$) of the electron, the path of the electron becomes so much deviated that we can speak of a collision. Thus the collisional cross-section is

$$S = \pi x^2 = \pi\frac{Z^2e^4}{w^2} = \pi\frac{Z^2e^4}{\gamma'^2 W_e^2} = \frac{4\pi}{9\gamma'^2}\frac{Z^2e^4}{(kT_e)^2}. \tag{60}$$

As because of the neutrality of the gas $n_e/n = Z$, we obtain from (58) and (55)

$$\sigma_{\|} = \frac{3\sqrt{3}}{4\pi}\gamma'^2\frac{(kT_e)^{\frac{3}{2}}}{Ze^2\sqrt{m_e}}. \tag{61}$$

The factor γ' is of order unity.

The conductivity of a completely ionized gas has been calculated by exact methods, first by Chapman (1928) and later by Cowling (1945). According to their results γ' varies slowly with temperature, density, and ionization. As an average, which in any case gives the order of magnitude, we may put

$$\gamma' = 0{\cdot}4, \tag{62}$$

which gives

$$S = 8{\cdot}7\frac{e^4}{k^2}\frac{Z^2}{T_e^2} = 2{\cdot}4.10^{-5}\frac{Z^2}{T_e^2}\text{ cm.}^2 \tag{63}$$

and

$$\sigma_{\|} = 1{\cdot}4.10^7 T_e^{\frac{3}{2}}/Z. \tag{64}$$

As, according to (54), (57), and 3.21 (39), we have

$$\omega_e\tau_e = ec^{-1}H(3km_e)^{-\frac{1}{2}}T_e^{-\frac{1}{2}}(nS)^{-1} = 10^6HT_e^{\frac{3}{2}}n_e^{-1}Z^{-1}, \tag{65}$$

the cross-conductivity is

$$\sigma_{\perp} = \sigma_{\|}[1+\omega_e^2\tau_e^2]^{-1} = [0{\cdot}7.10^{-7}ZT_e^{-\frac{3}{2}}+0{\cdot}7.10^5H^2n_e^{-2}T_e^{\frac{3}{2}}Z^{-1}]^{-1}. \tag{66}$$

It is of interest to compare the cross-section of ions and of molecules. For $T_e = 6{,}000°$ and $Z = 1$ we obtain from (63) $S = 0{\cdot}7.10^{-12}$ cm.^{2}, which is about a thousand times more than in the case of neutral atoms (molecules). This means that already when the degree of ionization has reached 10^{-3}, the interaction between electrons and ions in the photosphere begins to exceed the interaction between electrons and molecules. In Fig. 3.2 *c* the resistance X_e^i becomes smaller than X_e^M, and as X_i^M is much smaller than both of these, the energy is transferred from the electrons to the gas through X_e^i. Thus as soon as the degree of ionization exceeds about 10^{-3}, X stands for X_e^i in Fig. 3.2 *d*, and S refers to the cross-section of the ions. The molecules do not appreciably affect the conductivity.

The conductivity formula for a slightly ionized gas without magnetic field has been checked experimentally (see Loeb, 1939, p. 188). For a completely ionized gas no experimental check has been made. Nor is there any confirmation of the formula for the conductivity in a magnetic field.

Finally it must be observed that σ is not constant when the density of charged particles is affected by the discharge. In a self-sustained discharge n_e is usually proportional to the current density i. This means that σ is inversely proportional to i and the electric field E independent of i.

3.24. *Conductivity in cosmic physics.* Putting $Z = 1$ in (64) we obtain $\sigma_{\parallel} = 1{\cdot}4.10^{13}$ e.s.u. for $T = 10{,}000^\circ$, and $\sigma_{\parallel} = 1{\cdot}4.10^{16}$ e.s.u. for $T = 10^6$ degrees.

As a comparison the conductivity of copper is 5.10^{17} e.s.u. Hence the conductivity of a completely ionized gas is not far below that of metallic conductors. According to § 3.12 the conductivity is transformed as η^{-1}. This means that if we wish to construct a scale model with linear dimensions reduced by a factor of, say, 10^{-10}, we ought to make it of a substance with a conductivity which is many orders of magnitude higher than that of copper in order to represent a completely ionized gas of cosmic dimensions. This shows that the parallel conductivity in cosmic physics must often be considered as extremely good. In fact, in most problems concerning highly ionized gases we could consider it as infinite. This means that Ohm's law is of little importance in cosmic physics: we cannot find a current by calculating the voltage and resistance of a circuit, because we have seldom to deal with stationary currents. Instead the inductance enters as a dominating factor. In this respect cosmic electrodynamics is more related to high-frequency technics than to direct current problems. It must be observed, however, that cosmic problems derive a quite special character from two circumstances: most conductors are fluids, and strong magnetic fields are usually present. Some consequences of this will be discussed in § 3.41 and in § 4.1.

What has been said above does not altogether hold for problems concerning the environment of the earth. Here the dimensions are too small and the temperature, the degree of ionization, and hence the conductivity too low.

3.25. *Diffusion.* The diffusion of charged particles causes a current as soon as the density n_k has a gradient (see e.g. Cobine, 1941, p. 53). This current i can be calculated if, in the formulae of the preceding paragraphs, the electric field E is replaced by $E-(kT_k/e_k n_k)\operatorname{grad} n_k$. For example, the electronic current is given by

$$i_e = \sigma_e\left(E-\frac{kT_e}{e_e n_e}\operatorname{grad} n_e\right), \tag{67}$$

where e_e (< 0) is the electronic charge and σ_e the electronic conductivity which is given by 3.2 (29) or in the case of a magnetic field by 3.21 (40). For the ionic diffusion we obtain a similar expression:

$$i_i = \sigma_i\left(E-\frac{kT_i}{e_i n_i}\operatorname{grad} n_i\right), \tag{68}$$

where T_i is the ionic temperature, and σ_i is similar to σ_e. Note that $e_i > 0$.

Suppose that, in the absence of an electric field, a gas is ionized within a certain volume within which there are n_e electrons and n_i ions per unit volume. Initially we have $n_e = n_i$. At the border of the ionized region electrons and ions diffuse outwards. Because of the higher mobility of the electrons, these diffuse more rapidly, so after some time there is immediately outside the border a region containing a surplus of electrons, whereas immediately inside the border there are more ions than electrons. Hence an electric field is produced which impedes the outward motion of the electrons and accelerates the ionic diffusion. As, in the cases of interest in cosmic physics, the relative difference between n_e and n_i can never be large (see § 1.4), the electric field soon attains such a value that the diffusion of the electrons equals that of the ions. This type of diffusion is called *ambipolar* diffusion. It is a very important phenomenon in ordinary gaseous discharges (see Engel and Steenbeck, **1**, 1932, p. 197, or Cobine, 1941, p. 48).

The ambipolar diffusion is easily treated by (67) and (68). Putting $n_i \approx n_e = n$; $i_e + i_i = 0$ and introducing $|e|$ for the absolute value of the electronic charge, we find the electric field E produced by the difference in diffusion of electrons and ions:

$$E = -\frac{\sigma_e T_e - \sigma_i T_i}{\sigma_e + \sigma_i} \frac{k}{|e|n} \operatorname{grad} n. \tag{69}$$

The rate of diffusion corresponds to the current components

$$i_e = -i_i = \frac{\sigma_e \sigma_i}{\sigma_e + \sigma_i} \frac{k(T_e + T_i)}{|e|n} \operatorname{grad} n. \tag{70}$$

If $\sigma_i \ll \sigma_e$ we have approximately

$$i_e = -i_i = \frac{\sigma_i k(T_e + T_i)}{|e|n} \operatorname{grad} n. \tag{71}$$

3.3. Diamagnetism of an ionized gas

According to formula 2.2 (5) a charged particle which moves in a magnetic field H and has the energy $W_\perp$ perpendicular to the magnetic field has a magnetic moment μ:

$$\mu = \frac{W_\perp}{H}. \tag{1}$$

The condition for this is that the mean free path λ is much larger than the radius of curvature ρ. The moment is directed antiparallel to the magnetic field, so that the spiralling particle is diamagnetic.

It is not legitimate, however, to conclude from this that a gas consisting of charged particles is diamagnetic. From the classical theory of magnetic properties of metal conductors it is well known that no resultant diamagnetism is caused by the conduction electrons in a metal. As shown by Bohr, this is due to the fact that electrons are reflected against the walls of the conductor. These electrons give rise to currents neutralizing the currents of the electrons in the interior so that the diamagnetism cancels. The effect is shown in Fig. 3.3. Electrons are uniformly distributed over the whole cross-section of the conductor, and at every point the resultant current is zero, because of the isotropic distribution of the spiralling electrons. Even near the wall the electrons are isotropically distributed because of the perfect reflection of the walls.

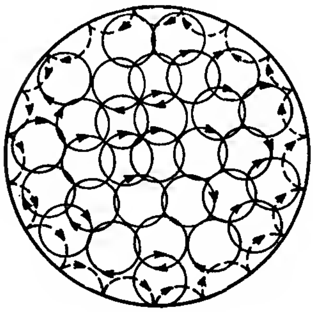

FIG. 3.3. An electronic gas enclosed by reflecting walls is non-magnetic because the diamagnetism of the electrons in the interior is compensated by electrons reflected at the walls.

As the isotropy is a consequence of thermodynamic equilibrium, it is evident that an ionized gas in such a state cannot be diamagnetic. This has been demonstrated by Cowling (1929) in connexion with a discussion on the radial limitation of the solar magnetic field.

On the other hand, as a single spiralling particle produces a diamagnetic moment, it seems reasonable that a gas consisting of an aggregate of such particles could be diamagnetic when it is *not* in thermodynamic equilibrium. The importance of this is evident in view of the fact that discharges are in a state very far from equilibrium. Let us discuss a simple case which shows how a gas may become diamagnetic.

Suppose that a cylindrical wall in a homogeneous magnetic field parallel to the cylinder axis encloses a gas which consists of n electrons per unit volume (see Fig. 3.4), and that the wall is perfectly reflecting and the gas in thermodynamic equilibrium and hence non-magnetic. Suppose further that the density is so small that $\omega\tau \gg 1$, which according to 3.21 (43) means that the electrons spiral many turns before they collide.

The pressure in the gas is nkT (T = temperature). This is also the force which a unit surface of the wall exerts on the gas. Let ν be the number of electrons which hit this surface per unit time. The average force per electron is given by $f = nkT/\nu$, and this force makes the ν electrons drift with the velocity $u_\perp$ which equals cf/eH [compare 2.2

(23)]. Hence the pressure of the wall gives a current

$$I_W = veu_\perp = cvf/H = cnkT/H. \tag{2}$$

This is the current per unit length of the cylinder due to electrons reflected at the wall. Denoting the cross-section of the cylinder by S and supposing the radius of curvature of the electronic paths to be small compared with $S^{\frac{1}{2}}$, the magnetic moment of the wall current i is

$$M_W = SI_W/c = SnkT/H. \tag{3}$$

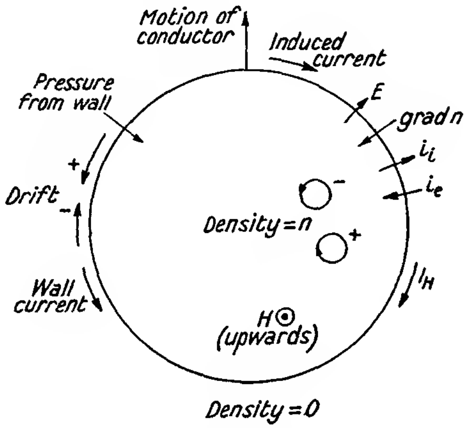

FIG. 3.4. Electronic gas with density n within a cylinder. The density gradient at the wall produces a drift ('wall current') which exactly compensates the Hall current, so that no magnetic effect occurs. If the wall is taken away the wall current disappears and the uncompensated Hall current makes the gas diamagnetic. In the absence of walls the expansion of the gas induces a current.

As the number of electrons per unit length of the cylinder is Sn, and each of these on an average has a moment $-W_\perp/H$ which equals $-kT/H$, their resulting moment is

$$M = -SnkT/H. \tag{4}$$

Hence the moment of the wall current exactly compensates the moment of the interior electrons, as expected.

Let us now suddenly take away the wall and the reflected electrons. This means that the wall current disappears, but it does not immediately disturb the motion of the other electrons. Certainly in the absence of a wall the electrons will diffuse outwards, but at sufficiently low pressures this is a very slow process due to the magnetic field. In fact the electrons continue to spiral in the same orbits as before until they accidentally collide. Hence at least the immediate result is that the gas becomes diamagnetic with a moment M_1 per unit volume.

$$M_1 = -n\mu = -nkT/H. \tag{5}$$

As in the interior of the gas the electronic motion is still isotropic, the diamagnetism may be considered to be produced by the electrons at the border. As the reflected electrons have been taken away, the motion near the border is anisotropic.

Looking at the problem from a macroscopic point of view the border current may be considered as a product of the density gradient and, as will now be shown, it can be interpreted as a Hall current. According to § 3.25 an electronic density gradient grad n gives the same electric current as a field $E = -(kT/en)\text{grad}\, n$. Hence the Hall current is

$$i_H = -\sigma_H(kT/en)\text{grad}\, n. \tag{6}$$

Our formulae for the conductivity do not refer to a pure electronic gas but to a gas which may consist of neutral molecules, electrons, and ions. We consider in a preliminary way such a gas and shall later return to the electron gas by making $\omega\tau$ very large, which means that the influence of the other constituents vanishes. Inserting 3.21 (41) and 3.21 (40) into (6) we find the current density

$$i_H = -\frac{e\omega\tau^2}{m(1+\omega^2\tau^2)}kT\,\text{grad}\, n, \tag{7}$$

or, because of 3.21 (39),

$$i_H = c\frac{kT}{H}\frac{\omega^2\tau^2}{1+\omega^2\tau^2}\text{grad}\, n. \tag{8}$$

If in a plane perpendicular to the magnetic field the density is constant within a certain region and zero outside it, there is a current I_H at the border. This current can be found by integrating (8):

$$I_H = c\frac{nkT}{H}\frac{\omega^2\tau^2}{1+\omega^2\tau^2}. \tag{9}$$

A surface S encircled by a current I_H has a magnetic moment

$$M\ (= -SI_H/c).$$

That the sign is negative is easily seen from Fig. 3.4. The moment, M_1 per unit volume is given by

$$M_1 = -\frac{nkT}{H}\frac{\omega^2\tau^2}{1+\omega^2\tau^2}. \tag{10}$$

For $\omega\tau \gg 1$ this result agrees with (5).

When $\omega\tau$ is small the magnetic moment of a single electron is no longer given by (1), because when deriving this in § 2.2 we have supposed that the particle made a large number of turns between two collisions. In this case the derivation of the magnetic moment from the Hall current is preferable, because it is more general.

The existence of a 'Hall diffusion current' is an indication of lack of equilibrium. Hence the occurrence of diamagnetism as a product of such a current is not in conflict with Bohr's theorem.

Except in the case $\omega\tau \gg 1$ the gas will spread more and more. Suppose that we try to prevent this by applying at the border an electric field which makes all those electrons turn back which attempt to escape. This electric field produces a Hall current which compensates the Hall diffusion current and hence makes the gas non-magnetic as if enclosed by a reflecting wall.

That an ionized gas in the absence of walls is diamagnetic can also be shown in the following way. Consider the case of an electron gas in a homogeneous magnetic field discussed above. When the wall is taken away, the gas would expand immediately if the magnetic field were not present. The magnetic field impedes the expansion and hence a force from the field acts upon the gas. This force can be considered as due (see Fig. 3.4) either to a motional induced current at the border of the gas or, which is only another way of expressing the same thing, to a difference in magnetostatic pressure inside and outside the gas. As the magnetostatic pressure in a body with the susceptibility κ is

$$(1+4\pi\kappa)H^2/4\pi$$

and the difference between this pressure and the magnetostatic pressure outside the body must equal the gas pressure nkT, we obtain

$$-\kappa H^2 = nkT,$$

which, as $\kappa H = M_1$, gives the same result as above. When neutral molecules are present also, the diamagnetism becomes smaller, because the magnetic pressure need only take up the difference in diffusion with and without magnetic field.

Our discussion of an electron gas is of interest because it shows that under certain conditions a charged particle gas may be diamagnetic. In cosmic physics a gas always contains about the same number of positive and negative particles. We shall now try to analyse the magnetic properties of such a gas.

In § 3.25 we have seen that when in an ionized gas the density of charged particles is great, the diffusion becomes ambipolar, i.e. an electric field is produced which makes the electronic diffusion decrease and the ionic diffusion increase so that both reach the same value. The drift of ions and electrons produced by this electric field changes the Hall current. If we add the Hall currents from the electrons and from the ions, we have

$$i_H = \omega_e \tau_e i_e - \omega_i \tau_i i_i, \tag{11}$$

Inserting 3.25 (70) we find

$$i_H = (\omega_e \tau_e + \omega_i \tau_i) \frac{\sigma_e \sigma_i}{\sigma_e + \sigma_i} \frac{k(T_e + T_i)}{|e|n} \operatorname{grad} n, \tag{11'}$$

where σ_e and σ_i mean the electronic and ionic conductivities perpendicular to the magnetic field [see 3.21 (40)]

$$\sigma_e = \frac{e^2 n \tau_e}{m_e(1+\omega_e^2 \tau_e^2)}, \tag{12}$$

$$\sigma_i = \frac{e^2 n \tau_i}{m_i(1+\omega_i^2 \tau_i^2)}, \tag{12'}$$

and [see 3.21 (39)]

$$\omega_e = |e|H/m_e c, \tag{13}$$

$$\omega_i = |e|H/m_i c. \tag{14}$$

If as above we integrate and put $\mu = -I_H/cn$, we find

$$\mu = \frac{k(T_e + T_i)}{H} \frac{\omega_e \tau_e + \omega_i \tau_i}{\omega_e \tau_e(1+Z_e) + \omega_i \tau_i(1+Z_i)}, \tag{15}$$

with

$$Z_e = (\omega_e \tau_e)^{-2}, \qquad Z_i = (\omega_i \tau_i)^{-2}. \tag{16}$$

At low pressures we can neglect Z_e and Z_i and find

$$\mu = \frac{k(T_e + T_i)}{H}, \tag{17}$$

which, because $W_\perp = kT$, is in agreement with (1) and represents the sum of the magnetic moments of one electron and one ion. For high pressures we have

$$\mu = \frac{k(T_e + T_i)}{H} \omega_e \tau_e \omega_i \tau_i = \frac{e^2 \tau_e \tau_i}{m_e m_i c^2} k(T_e + T_i) H. \tag{18}$$

Experimental data concerning the diamagnetism of ionized gases are scarce. Certainly since the time of Faraday it has been known that flames are diamagnetic. The repulsion of a flame by a magnet was studied a great deal during the nineteenth century, but quantitative results are lacking. No recent measurements seem to exist: the field is as completely out of fashion as the unipolar inductor and is not even mentioned in modern text-books. Reviews of the investigations are found in old handbooks, e.g. Graetz (1915). The observed effects are probably too large to be explained as due to molecular diamagnetism, so it is not unlikely that they are of the kind considered in this paragraph.

Although the discharge plasmas have been studied much more than flames, there seems to exist only one measurement of the diamagnetism of a plasma. This is made by Steenbeck (1936), who also attempts to

give a theory of it. His formula, which is similar to (10) but not identical with it, has been criticized by Tonks (1939), who gives a formula which in certain respects is similar to (15). As has been shown above, there is no doubt that formula (10), which refers to a pure electron gas, cannot be applied to a plasma, where, according to what is stated above, the diamagnetism is connected with the ambipolar diffusion. Steenbeck's experimental results confirm qualitatively his formula, but quantitatively his theoretical values of the diamagnetism seem to be about 15 times too large. The discrepancy is not definite because of the difficulty in measuring accurately some quantities, but his experimental results seem to be at least as well reconcilable with our formula (15). Compare also Rompe and Steenbeck (1939).

When studying the often enormous effect of a magnetic field on one moving particle we are tempted to conclude that even the effect on an ionized gas must be enormous. When the gas is at rest and in equilibrium this conclusion is illegitimate. Only a motion of the gas (eventually of the charged particle component of the gas) can produce currents and hence diamagnetic effect. The diamagnetism which we have considered is a product of the expansion of a gas, even if at low pressures and strong magnetic fields the expansion may be very slow. A diamagnetism of this kind is not very important in the laboratory but may be worth considering in cosmic physics. It is of special importance in an inhomogeneous magnetic field. Such a field tends to push a diamagnetic body into the weakest parts of the field. As an expanding gas is diamagnetic, this means that it prefers to expand in the direction where the magnetic field, which counteracts the expansion, is weakest.

3.4. Constriction of a discharge

A discharge in a gas may fill the whole space between the electrodes or be confined to a narrow channel. The former is the case for a glow discharge at low pressure. As an example of the latter type of discharge, which is called 'constricted', we may take a spark in air at atmospheric pressure. Constriction occurs in general at high pressures and strong currents.

There are some phenomena in the field of cosmic physics which may be interpreted as constricted discharges. A weak aurora is in general diffuse. In fact the most common auroral form is the diffuse arc. If its intensity increases it often 'dissolves' itself into auroral rays. It seems probable that this could be interpreted as a constriction of a discharge when the current exceeds a certain limit (see § 6.4). Another phenomenon

where we may have to do with a constricted discharge is the solar prominences (see § 5.6).

The constriction of a discharge is a very complicated phenomenon, because there are so many different factors which determine the constriction (see Engel–Steenbeck, **2**, p. 138, and Cobine, p. 317). Although it has been studied in detail for many special cases, a general survey of it seems to be lacking. If this is the state for the laboratory discharges, a discussion of the importance of the phenomenon in cosmic physics is certainly very difficult.

The constriction is often connected with a 'falling characteristic', by which we mean that the electric field necessary to maintain the discharge is a decreasing function of the current density. If the total discharge current is given, the field becomes smaller when the current concentrates in a small channel than when it fills the whole space. A discharge in general adjusts itself so that the field becomes a minimum. For example, in the cathode fall of a glow discharge the current density attains, if possible, that value for which the field has its smallest value. We are not actually concerned with this phenomenon here, but the conditions are somewhat similar in a plasma.

At atmospheric pressure discharges are diffuse at very low currents (e.g. in ionization chambers). At currents of the order of a few amperes or more, discharges become constricted. Examples are an arc, a spark, or a flash of lightning. The discharge channel is heated to several thousand degrees. The electronic temperature is usually almost the same as the gas temperature, and hence relatively low, and most of the ionization is due to temperature ionization (atomic collisions). The decrease in temperature towards the cool surrounding gas is very rapid.

The arc plasma does not obey the similarity laws of § 3.12. Instead it obeys rather complicated laws, the essence of which is that the heat produced by the electric current shall cover the thermal losses, mainly through convection and conduction, to the cold surroundings (see Cobine, p. 317). If we change a discharge according to the similarity laws of § 3.12, the conditions for thermal constriction become less favourable the lower the pressure (and hence the larger the dimensions). This is mainly due to the fact that the energy Ei developed per unit volume varies as η^{-3}, and hence becomes small when we go to larger dimensions and lower pressures. It is a general experience that at atmospheric pressure discharges are usually constricted, whereas at pressure below, say, 1 mm. Hg constriction becomes a more rare phenomenon.

It should be observed that even at low pressure constriction may occur

if certain gases (e.g. CO_2) are present. This constriction has nothing to do with the thermal effect discussed above. More detailed knowledge of its causes seems to be lacking.

Constriction may also be due to the electromagnetic attraction between parallel currents. Suppose that the discharge flows in a cylinder and that the current density $i(r)$ is a function of the distance r from the axis. The magnetic field $H(r)$ produced by the discharge is

$$H(r) = \frac{4\pi}{cr} \int\limits_0^r i(r)r\,dr. \tag{1}$$

A unit volume at distance r from the axis is acted upon by the force

$$\mathbf{f} = (1/c)[\mathbf{iH}], \tag{2}$$

or

$$f = \frac{4\pi i(r)}{c^2 r} \int\limits_0^r i(r)r\,dr = \frac{1}{4\pi}\left(\frac{H^2}{r} + H\frac{\partial H}{\partial r}\right), \tag{3}$$

which is directed towards the axis. This force causes a constriction of the discharge, because all charged particles drift towards the axis. The effect is known as 'pinch effect' (Tonks, 1937; Dow, 1944, p. 434). The pressure at the axis becomes higher than in the surrounding gas. In fact we have

$$\frac{\partial p}{\partial r} = -f, \tag{4}$$

or from (3) after integration

$$p = p_\infty - \frac{1}{8\pi}H^2 + \frac{1}{4\pi}\int\limits_r^\infty \frac{H^2}{r}\,dr. \tag{5}$$

3.41. *Application to cosmic physics.* We have seen that the constriction of a laboratory discharge is due in some cases to the heating of the gas through the discharge, in other cases to electromagnetic attraction. The heat produced per unit volume is proportional to iE, and hence should be transformed as $\eta^{-2}\eta^{-1} = \eta^{-3}$ (compare 3.14). The force f ($= iH/c$) causing electromagnetic constriction is also transformed as η^{-3}. The former effect gives rise to an increase in temperature, which must have a certain absolute value in order to be of importance. The latter effect, however, should be compared with other forces, the most important one being the force f ($= qE$) of the electric field E upon a space-charge q. This force is also transformed as η^{-3}. Thus the relative importance of electromagnetic constriction does not diminish as does the heat constriction. Consequently it seems likely that electromagnetic constriction constitutes the most important effect in cosmic

physics. This does not at all mean that heat effect could be neglected—owing to the enormous currents in cosmic physics they are probably also important. Nor does it mean that other causes for constriction are ruled out.

Hence for the application to cosmic physics we ought to study the constriction due to magnetic effects. This is certainly a complicated phenomenon, and only some hints about factors of importance can be given here.

Let us discuss the simple case of an ionized gas in a homogeneous electric field E in the absence of an imposed magnetic field. A current with constant density i is produced:

$$i = \sigma E. \tag{6}$$

Here σ means the conductivity, which is given by 3.2 (29) as long as the magnetic field is zero. When the electric field is strong enough to produce a current which gives rise to an appreciable magnetic field H, the conductivity is given by 3.21 (40) instead. Putting

$$\alpha^2 = \frac{\omega_e^2 \tau_e^2}{H^2} = \frac{e^2 \tau_e^2}{m_e^2 c^2}, \tag{7}$$

we obtain for the conductivity

$$\sigma = \frac{\sigma_\parallel}{1+\alpha^2 H^2}. \tag{8}$$

The magnetic field depends upon the boundary conditions. If the conditions are symmetrical with respect to the z-axis, supposed to be parallel to E, we have

$$H = \frac{4\pi}{cr} \int\limits_0^r ir\, dr, \tag{9}$$

where r is the distance from the z-axis. Differentiating (9) and inserting (6) and (8) we obtain

$$\frac{dH}{dr} = \frac{\beta}{1+\alpha^2 H^2} - \frac{H}{r}, \tag{10}$$

with

$$\beta = \frac{4\pi\sigma_\parallel E}{c}. \tag{11}$$

We have found that the magnetic field of the current causes a concentration of the discharge to the symmetry line. The problem is, however, more complicated than this, because the electric field and the current magnetic field produce together a drift of charged particles. This drift is directed towards the axis and causes a continuous increase in charged particle density near the axis until compensating effects occur. The most important of these is the outward diffusion due to the

density gradient dn_e/dr, which according to 3.25 (67) causes a current $-\sigma_\perp kT_e \operatorname{grad} n_e/e_e n_e$ in the radial direction and a Hall current in the z-direction, which according to 3.21 (41) is $-\omega_e \tau_e$ times the radial current. Adding these currents to those caused directly by the electric field we obtain

$$i_z = \sigma_\perp \left(E_z + \frac{kT_e}{e_e n_e} \frac{dn_e}{dr} \omega_e \tau_e \right), \tag{12}$$

$$i_r = \sigma_\perp \left(-E_z \omega_e \tau_e - \frac{kT_e}{e_e n_e} \frac{dn_e}{dr} \right). \tag{13}$$

If equilibrium is reached we have

$$i_r = 0.$$

This gives from (13) $$\frac{1}{n_e} \frac{dn_e}{dr} = -\frac{e_e E_z}{kT_e} \omega_e \tau_e. \tag{14}$$

Inserting this value into (12) we find

$$i_z = \sigma_\perp (1 + \omega_e^2 \tau_e^2) E_z = \sigma_\parallel E_z, \tag{15}$$

a result which has already been obtained in § 3.21, p. 50.

If 3.21 (39) and (9) are introduced into (14) we may calculate n_e, and hence i, as a function of r in a stationary state. The result is that the current density is almost constant near the axis but decreases as r^{-3} for large values of r. Hence a constriction occurs.

In a completely ionized gas the conditions are somewhat different because according to 3.23 (61) the conductivity is independent of the density. Hence we should expect a uniform current density. The pressure of the gas increases towards the z-axis.

This stationary state is probably never attained in cosmic physics. The reason for this can be understood from the following discussion.

Consider a cylinder (axis coinciding with the z-axis, surface given by $r = r_0$) consisting of a solid conductor. If an alternating electric field E is applied in the z-direction, the current is confined to a thin layer at the surface because of the skin effect (see e.g. Harnwell, 1938, p. 315). In fact, the current penetrates to a depth Δ of the order of

$$c(\omega\sigma)^{-\frac{1}{2}},$$

where ω is the frequency of the electric field, σ the conductivity, and c the velocity of light. If instead a constant field is suddenly applied, the current starts at the surface but at the first moment no current flows in the interior of the conductor. The current-carrying layer gradually

increases, so that after the time τ it has a thickness of the order

$$\Delta = c(\tau/\sigma)^{\frac{1}{2}}.$$

In cosmic physics a completely ionized gas—such as we have in the solar corona and in parts of interstellar space—has a conductivity of the order of, say, $\sigma = 10^{13}$–10^{16} e.s.u. If a solid conductor had a conductivity of 10^{13} e.s.u., the current sheath would after one year (3.10^7 sec.) have a thickness of 5.10^7 cm., which for the problems mentioned above is next to nothing. Consequently in a solid conductor of cosmic dimensions an applied electric field cannot cause a current within a reasonable time except in a very thin surface layer. As in high-frequency technics, it is easier to produce a current in a dielectric than in the interior of a conductor.

If we pass from a solid conductor to the ionized gas, the skin effect changes character. Suppose that a gas is ionized only within a cylinder, limited by the surface $r = r_0$ and two circular electrodes $z = \pm z_0$. When an electric field E_z is suddenly applied between the electrodes, current starts at the surface of the cylinder. Due to the attraction between parallel currents, a force is produced directed towards the axis. The gas begins to stream inward with the velocity v. An induced field $\mathbf{E}' = [\mathbf{vH}]c^{-1}$ is produced which reduces the applied field. The result is probably that current flows only near the z-axis.

This 'inverse skin effect' has never been studied but may be very important in cosmic physics. In order to discuss it mathematically we start with Maxwell's equations:

$$\text{curl}\,\mathbf{H} = \frac{4\pi}{c}\mathbf{i}, \tag{16}$$

$$\text{curl}\,\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{H}}{\partial t}, \tag{17}$$

where the displacement current has been neglected and the permeability put equal to unity. Further we have

$$\mathbf{i} = \sigma\{\mathbf{E}+c^{-1}[\mathbf{vH}]\}. \tag{18}$$

If the gas has the density ρ and the pressure is p, the force equation gives

$$\rho\frac{d\mathbf{v}}{dt} = \frac{1}{c}[\mathbf{iH}]-\text{grad}\,p. \tag{19}$$

Further we have the continuity equation

$$\rho\,\text{div}\,\mathbf{v}+\frac{d\rho}{dt} = 0. \tag{20}$$

A sixth equation for the determination of the six variables E, H, i, v, ρ, and p is supplied by the gas equation

$$\rho = \frac{m}{kT}p, \tag{21}$$

where m is the average molecular mass, k Boltzmann's constant, and T the temperature. The latter may be considered as a parameter, or, rather, to be given by the condition that the energy developed by the discharge should equalize the energy losses. The system of equations is similar to that which will be treated in § 4.2, where magneto-hydrodynamic waves are deduced from it. A difference is that in the present case the medium is supposed to be compressible. The boundary conditions are also different.

An exact solution of the system of equations encounters mathematical difficulties. An approximate solution which seems to give the essential features of the phenomenon may be obtained from the three first equations alone in the following way.

In order to simplify the calculations, let us confine ourselves to the plane case, the variables being functions of x only, the y–z plane being a plane of symmetry. The only components whose values are not zero are E_z, H_y, i_z, v_x (and ρ and p). Hence equations (16), (17), and (18) become

$$\frac{\partial H}{\partial x} = \frac{4\pi}{c}i, \tag{22}$$

$$\frac{\partial E}{\partial x} = \frac{1}{c}\frac{\partial H}{\partial t}, \tag{23}$$

$$i = \sigma\left(E + \frac{v}{c}H\right), \tag{24}$$

where the subscripts are omitted. Eliminating i from (22) and (24), differentiating, and introducing (23) we obtain

$$\frac{\partial^2 H}{\partial x^2} = \frac{4\pi\sigma}{c^2}\left(\frac{\partial H}{\partial t} + v\frac{\partial H}{\partial x} + H\frac{\partial v}{\partial x}\right). \tag{25}$$

As, because of the symmetry, $v = 0$ for $x = 0$, we put simply

$$v = -x/\tau_0, \tag{26}$$

where τ_0 is a constant. Then the equations are satisfied by

$$i = i_0\, e^{t/\tau_0} e^{-(x/x_0)^2}, \tag{27}$$

$$H = -\frac{4\pi i_0}{c} e^{t/\tau_0} \int_0^x e^{-(x/x_0)^2}\, dx, \tag{28}$$

$$E = e^{t/\tau_0}\left\{\frac{4\pi i_0}{c^2\tau_0} x \int_0^x e^{-(x/x_0)^2}\, dx + \frac{i_0}{\sigma} e^{-(x/x_0)^2}\right\}, \tag{29}$$

where i_0 is a constant and

$$x_0 = \sqrt{(\tau_0 c^2/2\pi\sigma)}. \tag{30}$$

The velocity v is independent of time, but i, H, and E increase by a factor e during the time τ_0. The breadth x_0 of the current is given by (30).

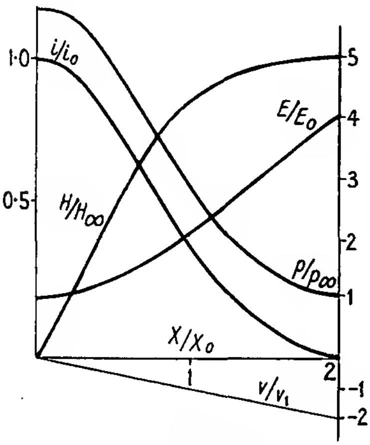

FIG. 3.5. 'Reversed skin effect.'

Introducing (26)–(28) into (19), (20), and (21) we may compute ρ and p. From (20) we find that ρ depends upon t as e^{t/τ_0}. In problems of interest in cosmic physics $\rho\, dv/dt$ in (19) is usually negligible. Hence we obtain

$$p = \frac{1}{c}\int ih\, dx,$$

showing that p varies as e^{2t/τ_0}. This is compatible with (21) if T is proportional to e^{t/τ_0}. In this case our solution would satisfy all the equations. If T varies in a more complicated way, as it can usually be expected to do, our solution is only approximate, but probably gives the essential features of the phenomenon.

The functions i, H, E, and p are plotted in Fig. 3.5. The electric field consists of the applied field which is constant and the field induced by the variation of H. The latter may be considered as due to the self-inductance of the circuit, which only allows the current to increase slowly. The whole phenomenon is related to the skin effect, which only permits a current to penetrate into a conductor gradually. The skin effect is complicated by the fact that our conductor is movable. In fact, the inward motion prohibits the current from spreading in the conductor.

In an ordinary (non-ionized) gas an electric field may produce a breakdown, characterized by the fact that the whole current flows in a small, highly ionized channel, where the density is small. We have found that in an ionized gas a somewhat similar phenomenon may occur. The whole current flows in a small channel; the rest of the gas can, because of a sort

of skin effect, carry no current. This is important in the theory of solar prominences (§ 5.62).

A careful analysis of the problem would be very desirable.

3.5. Maximum current density in an ionized gas

There seems to be an upper limit to the current density which a plasma is able to carry. This limit is much higher than the current densities of

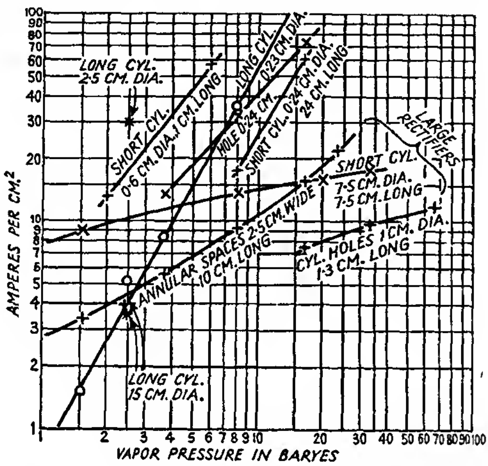

FIG. 3.6. Maximum current density in mercury gas at different pressures. (From Tonks, 1937.)

normal glow discharges. As the limit decreases with decreasing pressure, it is reached only in heavy arc discharges at very low pressures.

In a mercury rectifier the discharge takes place between a mercury cathode and an iron or graphite anode. The gas is mercury vapour at a pressure which depends upon the wall temperature but usually is in the range of 10^{-3}–10^{-1} mm. Hg. In large rectifiers the arc current may be many thousand amperes. The voltage between the electrodes is of the order 20 volts. If the arc current is increased continuously the rectifier behaves normally up to a certain limit, at which the voltage rises extremely rapidly to a thousand volts or more. Usually the rise is discontinuous so that the arc is extinguished.

Although the disrupture of an 'overloaded' rectifier is well known to all manufacturers of such devices, the literature on the field is scarce. An interesting survey has been given by Tonks (1937). A diagram in his paper shows for different devices the limiting current density as a function of the mercury pressure (Fig. 3.6). Although the values differ

considerably owing to the different experimental arrangements, the diagram indicates that the limiting current increases with pressure, and possibly is directly proportional to it.

Tonks discusses the possible causes of the current limitation. The pressure in the discharge is the sum of the electronic pressure, the ionic pressure, and the pressure of the neutral mercury gas. As the total pressure is given, the increase in electronic pressure, which accompanies an increase in current, leads to a rarefaction of the gas in the arc, which may stop the discharge. Another possible cause is the 'pinch effect' (see § 3.4). The electromagnetic attraction between parallel currents causes the arc to contract, and according to Tonks the contracting force surpasses all counteracting effects when the current exceeds a certain limit.

It seems likely that disrupture at high currents should occur also in other gases than mercury vapour. As completely ionized hydrogen is the main constituent of e.g. the solar corona, discharges in such a gas are especially important in cosmic physics.

When the ionization is due to a high temperature and not produced by the discharge, as in the case studied by Tonks, a disrupture may be caused by the following mechanism.

According to equation 3.2 (29), the current i is given by

$$i = \sigma E = \frac{e^2}{m_e} n_e \tau_e E. \tag{1}$$

An electron gains energy at the rate $(dW/dt)_+$ (compare 3.2 (10) and 3.2 (23)),

$$\left(\frac{dW}{dt}\right)_+ = eEu_e = \frac{e^2}{m_e}\tau_e E^2 = \frac{m_e}{n_e^2 e^2 \tau_e} i^2. \tag{2}$$

It loses energy at the rate $(dW/dt)_-$ (compare 3.2 (11) and 3.2 (14)),

$$\left(\frac{dW}{dt}\right)_- = \frac{\gamma_1 m_e}{M\tau_e}(W_e - W_M). \tag{3}$$

If the electronic temperature is T_e and the gas temperature T_M, the pressure

$$p = n_e k(T_e + T_M/Z), \tag{4}$$

where Z is the charge of the ions.

In a stationary state $(dW/dt)_+$ must equal $(dW/dt)_-$. With the help of (4), and observing that $W_e = \frac{3}{2}kT_e$; $W_M = \frac{3}{2}kT_M$, we obtain from (2) and (3):

$$i = \sqrt{\left(\frac{3\gamma_1}{2kMT_M}\right)} epf(x), \tag{5}$$

with
$$f(x) = \frac{\sqrt{(x-1)}}{x+1/Z} \tag{6}$$
and
$$x = T_e/T_M. \tag{7}$$
The electric field according to (1) is given by
$$E = \frac{m_e}{n_e \tau_e e^2} i, \tag{8}$$

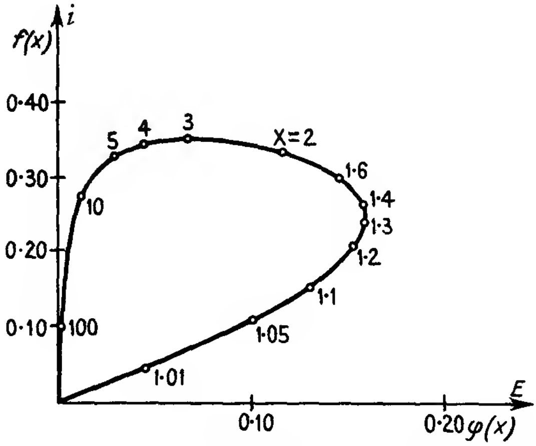

FIG. 3.7. Current voltage diagram of completely ionized gas.

or as $n_e \tau_e = Zn\lambda_e/v_e = Z/Sv_e = 9\gamma'^2 m_e^{1/2}(kT)^{3/2}/4\pi\sqrt{3}Ze^4$ (compare 3.23 (57) and 3.23 (60)),
$$E = \frac{4\pi\sqrt{3}}{9\gamma'^2}\frac{Ze^2\sqrt{m_e}}{(kT_e)^{3/2}} i, \tag{9}$$
$$E = \beta \frac{e^3\sqrt{m_e}}{k^2\sqrt{M}} \frac{p}{T_M^2} \varphi(x), \tag{10}$$
with
$$\beta = \frac{2\sqrt{2}\,\pi\sqrt{\gamma_1}\,Z}{3\gamma'^2} = 30Z, \tag{11}$$
$$\varphi(x) = \frac{1}{x^{3/2}} \frac{\sqrt{(x-1)}}{x+1/Z}. \tag{12}$$

Confining ourselves to the case of completely ionized hydrogen ($Z = 1$; $M = 1{\cdot}67 . 10^{-24}$ g.) we may calculate E and i as functions of x. Fig. 3.7 shows a diagram with E and i as axes. The field reaches its maximum for $x = 1{\cdot}32$ which gives $\varphi = 0{\cdot}16$. The current has its maximum for $x = 3$; $f = \sqrt{2}/4 = 0{\cdot}35$. The maximum current is given by
$$i = 0{\cdot}35(\tfrac{3}{2}\gamma_1/kMT_m)^{1/2} ep = 2{\cdot}24 . 10^{10} p/\sqrt{T_m} \text{ e.s.u.}$$

The maximum field

$$= 30.0{\cdot}16 \frac{e^3\sqrt{m_e}}{k^2\sqrt{M}} \frac{p}{T_M^2} = 650 \frac{p}{T_M^2} \text{ e.s.u.}$$

If either of these values is exceeded, no stationary state is possible and it seems likely that some sort of 'electron gas explosion' occurs. The temperature goes up and the density down unlimited. Our formulae, which refer to a stationary state, do not hold.

REFERENCES

General references: The theory of discharges in gases is treated in several modern text-books, e.g.

von Engel, A., and Steenbeck, M., 1932: *Elektrische Gasentladungen*, **1**, Berlin.

——— 1934: ibid. **2**, Berlin.

Loeb, L. B., 1939: *Fundamental Processes of Electrical Discharge in Gases*, New York.

Cobine, J. D., 1941: *Gaseous Conductors*, New York.

See also Compton, K. T., and Langmuir, I.: 'Electrical Discharges in Gases', *Rev. Mod. Phys.*, **2**, 124, 1930.

For the mathematical theory the reader is referred to: Chapman, S., and Cowling, T. G., 1939: *The Mathematical Theory of Non-uniform Gases*, Cambridge.

Birkeland, K., 1901: 'Videnskapsselskabets Skrifter I', *Mat.-nat. Kl.*, No. 1, Kristiania (Oslo).

Chapman, S., 1928: 'The Electrical Conductivity of Stellar Matter, *Mon. Not. R. Astr. Soc.*, **89**, 54.

—— and Cowling, T. G., 1939: *The Mathematical Theory of Non-uniform Gases*, Cambridge.

Cobine, J. D., 1941: *Gaseous Conductors*, New York.

Cowling, T. G., 1929: 'On the Radial Limitation of the Sun's Magnetic Field, *Mon. Not. R. Astr. Soc.*, **90**, 140.

—— 1945: 'Electrical Conductivity of an Ionized Gas in a Magnetic Field, with Applications to the Solar Atmosphere and the Ionosphere,' *Proc. Roy. Soc.* A **183**, 453.

Cravath, A. M., 1930: 'The Rate at which Ions lose Energy in Elastic Collisions', *Phys. Rev.*, **36**, 248.

Dow, W. G., 1944: *Fundamentals of Engineering Electronics*, New York.

von Engel, A., and Steenbeck, M., 1932: *Elektrische Gasentladungen*, **1**, Berlin.

——— 1934: ibid. **2**, Berlin.

Graetz, L., 1915: *Handbuch der Elektrizität und des Magnetismus*, Bd. **4**, 2, 780, Leipzig.

Harnwell, G. P., 1938: *Principles of Electricity and Electromagnetism*, New York.

Loeb, L. B., 1939: *Fundamental Processes of Electrical Discharge in Gases*, New York.

—— and Meek, J. M., 1941: *The Mechanism of the Electric Spark*, Stanford.

Rompe, R., and Steenbeck, M., 1939: 'Der Plasmazustand der Gase', *Erg. d. exakten Naturwiss.*, **18**, 257–377.

Steenbeck, M., 1936: 'Über die magnetischen Eigenschaften des Plasmas von Gasentladungen', *Wissenschaftliche Veröffentlichungen aus den Siemens-Werken*, **15**, 2, 1.

Compare also:

Tonks, L., 1939: *Phys. Rev.*, **56**, 360.

—— 1937: *Theory and Phenomena of High Current Densities in Low Pressure Arcs*. The Electrochemical Society. Preprint 72—13. St. Louis Meeting, Oct. 1937.

IV

MAGNETO-HYDRODYNAMIC WAVES

4.1. Introduction

If an electrically conducting medium is present in a magnetic field, any hydrodynamie motion will give rise to induced electrie fields which produce electric currents. Because of the magnetic field these currents will produce forces which change the state of motion. This coupling between mechanical and electromagnetic forces produces a type of wave motion, called *magneto-hydrodynamic waves* (Alfvén, 1942). As in cosmic physics magnetic fields are very important and matter frequently ionized, and hence conducting, magneto-hydrodynamic waves may be expected to be of considerable significance.

Let us, for the sake of simplicity, make the assumption that the magnetic field is homogeneous and has the strength H_0. We then lay a right-angled coordinate system with the z-axis parallel to the lines of magnetic force. We further assume that at a certain instant t_0 the whole fluid is at rest, with the exception of a pillar bounded in the xz-plane by $ABCD$ (Fig. 4.1) and extending indefinitely in the y-direction. This pillar moves with the velocity v in the direction of the y-axis.

A body moving in a magnetic field becomes electrically polarized in a direction at right angles to the magnetic field and to the direction of motion. The electric field thus produced is proportional to the vector product of the magnetic field and the velocity (compare 1.3):

$$\mathbf{E} = (\mu/c)[\mathbf{vH}_0].$$

In our case there consequently arises, in the moving part of the fluid, an electric field in the positive direction of the x-axis. The electromotive force induced in this manner produces currents, for we have assumed that the fluid is electrically conducting. The currents move in the direction of the induced electric field within $ABCD$, and are closed in the surrounding stationary fluid, so that the electric current system shown in Fig. 4.2 arises.

An electric current I in a magnetic field is acted upon by a force F which is at right angles to the current and to the magnetic field:

$$\mathbf{F} = \mu c^{-1}[\mathbf{IH}_0].$$

As the current in the moving part of the fluid flows in the positive direction of the x-axis, the force is negatively directed along the y-axis, i.e. it impedes the movement. In the adjacent parts of the stationary

fluid, however, the current flows in the negative direction of the x-axis, which causes the force to take the positive direction of the y-axis, so that it gives an acceleration in the direction of the initial movement. The layer of the fluid that was moving in the beginning is thus retarded

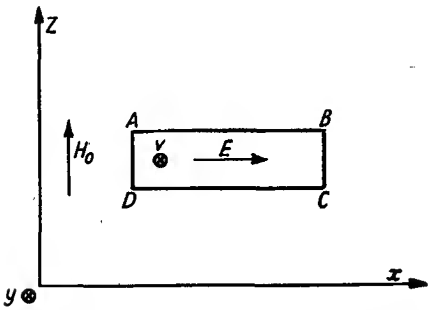

FIG. 4.1. When a pillar $ABCD$ moves in the y-direction, the magnetic field H_0 causes an electric field E.

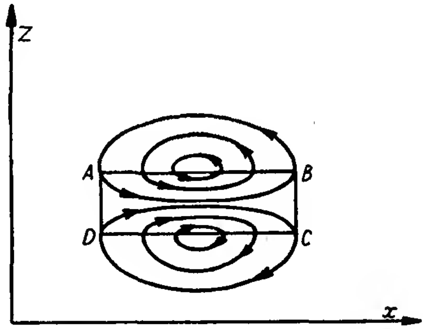

FIG. 4.2. Current system.

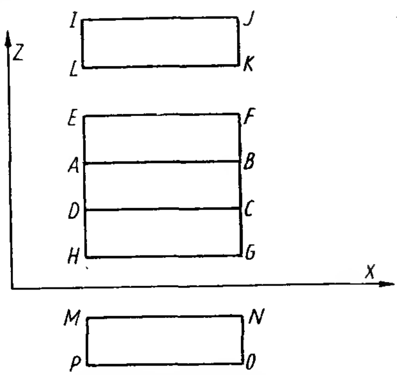

FIG. 4.3. Magneto-hydrodynamic waves transmit the state of motion of $ABCD$ to $EFBA$ and $DCGH$, and later to $IJKL$ and $MNOP$.

and the layers above and below are accelerated. *Thus the induced current system tends to transfer the initial movement to the surrounding layers.* The result will be that after some time $ABCD$ is at rest, while $EFBA$ and $DCGH$ (see Fig. 4.3) are moving in the direction of the y-axis. As the same mechanism is still acting, the state of motion will be transmitted in the positive and negative directions of the z-axis, so that after some time a part $IJKL$ and another $MNOP$ are moving, while the remaining parts of the fluid are at rest. In this way a magneto-

hydrodynamic wave is produced, which transmits the state of motion in the direction of the magnetic field.

The existence of magneto-hydrodynamic waves has not yet been demonstrated experimentally. In spite of the fact that they seem to constitute a simple corollary to well-known parts of classical physics, such a demonstration is highly desirable. A laboratory experiment with waves in mercury is feasible, but the low conductivity of mercury damps the waves very much. Because of the similarity principle (see § 3.12 and § 3.24) the equivalent conductivity in the cosmos is very much higher so the damping is less important than in a laboratory experiment.

4.2. Fundamental equations

In order to formulate the problem mathematically, we must start with Maxwell's equations

$$\operatorname{curl}\mathbf{H} = \frac{1}{c}\left(4\pi\mathbf{i}+\frac{\partial\mathbf{D}}{\partial t}\right), \tag{1}$$

$$\operatorname{curl}\mathbf{E} = -\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t}, \tag{2}$$

combined with

$$\mathbf{B} = \mu\mathbf{H}, \tag{3}$$

$$\mathbf{i} = \sigma\left\{\mathbf{E}+\left[\frac{\mathbf{v}}{c}\mathbf{B}\right]\right\}, \tag{4}$$

where σ is the electrical conductivity. We must add to these equations the hydrodynamic equation

$$\frac{d\mathbf{v}}{dt} = \mathbf{G}+\frac{1}{\rho}\left\{\left[\frac{\mathbf{i}}{c}\mathbf{B}\right]-\operatorname{grad}p\right\}, \tag{5}$$

where ρ means the mass density, p the pressure, and $\mathbf{G}$ the non-electromagnetic forces acting upon the fluid. The right side of the equation gives the forces acting upon a unit mass of the fluid.

4.3. Plane waves in incompressible fluid. Homogeneous field

We shall start by treating the simple case of plane waves in an incompressible fluid with infinite conductivity σ and constant density ρ. Then we have

$$\operatorname{div}\mathbf{v} = 0. \tag{1}$$

Further we assume that the primary magnetic field $\mathbf{H}_0$ is homogeneous and parallel to the z-axis of an orthogonal reference system, and that $\mathbf{G} = 0$. The displacement current $\partial\mathbf{D}/\partial t$ is negligible in comparison with the conduction current $\mathbf{i}$.

The magnetic field

$$\mathbf{H} = \mathbf{H}_0+\mathbf{h}, \tag{2}$$

where the field **h** is caused by the current **i**. In order to study a plane wave propagated in the direction of $\mathbf{H}_0$, we assume that all vectors are independent of x and y but depend upon z and the time t.

This implies, according to 4.2 (1) and 4.2 (2), that we have $i_z = 0$ and $H_z = \text{const.} = H_0$. Further, according to (1) we may put $v_z = 0$.

If we turn the coordinate system in such a way that $i_y = 0$, we obtain from 4.2 (1)

$$i_x = -\frac{c}{4\pi}\frac{\partial h_y}{\partial z}, \tag{3}$$

$$\begin{aligned} i_y &= i_z = 0, \\ h_x &= \text{const.} = 0, \\ H_z &= H_0. \end{aligned} \tag{4}$$

We introduce these values into 4.2 (5). As according to our assumptions $\operatorname{grad} p$ can have no components perpendicular to the z-axis, we obtain

$$\begin{aligned} \partial v_x/\partial t &= 0; \qquad v_x = \text{const.} = 0, \\ \frac{\partial v_y}{\partial t} &= \frac{\mu H_0}{4\pi\rho}\frac{\partial h_y}{\partial z}, \\ v_z &= 0, \end{aligned} \tag{5}$$

and further
$$\frac{\partial p}{\partial z} = -\frac{\mu}{8\pi}\frac{\partial(h_y^2)}{\partial z}. \tag{6}$$

Equation 4.2 (4) gives
$$\mathbf{E} = \frac{\mathbf{i}}{\sigma} - \mu\left[\frac{\mathbf{v}}{c}\mathbf{H}\right],$$

or, with (4) and (5),
$$\begin{aligned} E_x &= \frac{i_x}{\sigma} - \mu\frac{v_y}{c}H_0, \\ E_y &= E_z = 0. \end{aligned} \tag{7}$$

Equation 4.2 (2) gives
$$\frac{\partial h_y}{\partial t} = -\frac{c}{\mu}\frac{\partial E_x}{\partial z}. \tag{8}$$

From (7) and (8) we obtain

$$\frac{\partial^2 h_y}{\partial t^2} = H_0\frac{\partial^2 v_y}{\partial t\partial z} - \frac{c}{\mu\sigma}\frac{\partial^2 i_x}{\partial t\partial z}.$$

Introducing (3) and (5) we obtain

$$\frac{\partial^2 h_y}{\partial t^2} = \frac{\mu H_0^2}{4\pi\rho}\frac{\partial^2 h_y}{\partial z^2} + \frac{c^2}{4\pi\mu\sigma}\frac{\partial^3 h_y}{\partial^2 z\partial t}. \tag{9}$$

4.31. *Infinite conductivity.* In the case $\sigma = \infty$ we have

$$\frac{\partial^2 h_y}{\partial t^2} = \frac{\mu H_0^2}{4\pi\rho}\frac{\partial^2 h_y}{\partial z^2}. \tag{9'}$$

This is the wave equation and represents a wave with velocity V, where

$$V = \pm H_0\sqrt{(\mu/4\pi\rho)}. \tag{10}$$

The velocity of the electromagnetic-hydrodynamic wave is independent of the frequency as well as of the amplitude.

To find the order of magnitude of V, let us calculate the velocity for

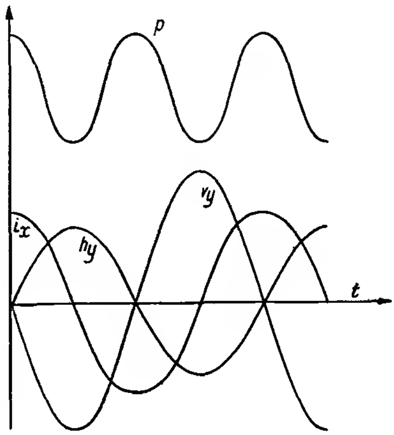

FIG. 4.4. Induced magnetic field h_y, velocity v_y, current i_x, and pressure P as functions of t in a sine wave.

the case $H_0 = 100$ gauss, $\mu = 1$, $\rho = 1$ g. cm.$^{-3}$ We find $V = 28$ cm./sec.

If we put

$$h_y = A\sin\omega(t-zV^{-1}), \tag{11}$$

we find

$$v_y = -\frac{A\sqrt{\mu}}{\sqrt{(4\pi\rho)}}\sin\omega\left(t-\frac{z}{V}\right), \tag{12}$$

$$i_x = A\frac{c\omega}{H_0\sqrt{}}\bigg/\left(\frac{\rho}{4\pi\mu}\right)\cos\omega\left(t-\frac{z}{V}\right), \tag{13}$$

$$E_x = A\frac{H_0\sqrt{\mu^3}}{c\sqrt{(4\pi\rho)}}\sin\omega\left(t-\frac{z}{V}\right), \tag{14}$$

$$p = p_0-\frac{\mu A^2}{8\pi}\sin^2\omega\left(t-\frac{z}{V}\right). \tag{15}$$

The magnetic lines of force, which with no waves were straight lines

$$x = x_0, \qquad y = y_0, \tag{16}$$

change their form as a consequence of the field h_y superimposed on H_0. As the angular coefficient of the lines of force must everywhere be given by

$$\frac{dy}{dz} = \frac{h_y}{H_z}, \tag{17}$$

we find, by introducing (11) into (17) and integrating, that the lines of force are sine curves with

$$x = x_0, \tag{18}$$

$$y = y_0 + \frac{A\sqrt{\mu}}{\omega\sqrt{(4\pi\rho)}}\cos\omega\left(t - \frac{z}{V}\right). \tag{19}$$

If we differentiate y with respect to t, we shall find that the magnetic

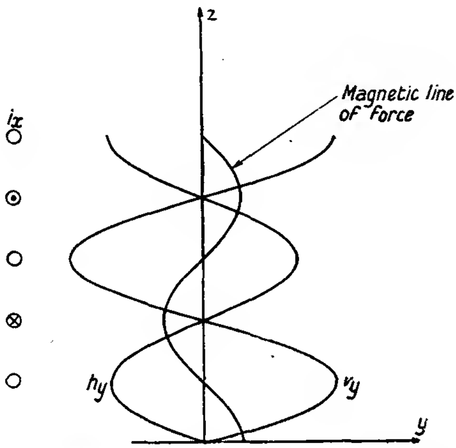

FIG. 4.5. Induced magnetic field h_y, velocity v_y, current i_x as functions of z at a certain moment. A magnetic line of force of the combined field H_0+h_y is also shown.

lines of force move with the same velocity, v_y [in accordance with (12)], as the fluid. The magnetic lines of force are 'frozen' into the liquid.

4.32. *Finite conductivity.* Suppose that σ is finite and h and v are parallel to the y-axis as earlier. Suppose further that h and v are functions of z and t only:

$$h = h_0\exp(\alpha z + j\omega t), \tag{20}$$

$$v = v_0\exp(\alpha z + j\omega t), \tag{21}$$

where $j = \sqrt{(-1)}$ and ω, α, h_0, and v_0 are constants, all of which except ω may be complex numbers.

As $\partial^2/\partial t^2 = -\omega^2$, $\partial^2/\partial z^2 = \alpha^2$, and $\partial^3/\partial z^2\partial t = j\omega\alpha^2$, we obtain from (9) with the help of (10):

$$\omega^2 + \left(V^2 + j\frac{c^2\omega}{4\pi\mu\sigma}\right)\alpha^2 = 0, \tag{22}$$

or

$$\alpha = \pm\frac{j\omega}{V}\left(1 + \frac{j\omega c^2}{4\pi\mu\sigma V^2}\right)^{-\frac{1}{2}}, \tag{23}$$

or for small damping approximately

$$\alpha = \pm\left(j\frac{\omega}{V}+\frac{\omega^2c^2}{8\pi\mu\sigma V^3}\right). \tag{24}$$

The distance z_0 in which the amplitude of the wave is reduced to $1/e$ is the inverse value of the real component of α. Thus we have

$$z_0 = \frac{8\pi\mu\sigma V^3}{\omega^2c^2} = \frac{2\mu\sigma V}{\pi c^2}\lambda^2 = \frac{\mu^{\frac{1}{2}}\sigma H_0}{\pi^{\frac{3}{2}}\rho^{\frac{1}{2}}c^2}\lambda^2 = \frac{\mu^{\frac{3}{2}}\sigma H_0^3}{\sqrt{\pi}\,c^2\rho^{\frac{3}{2}}\omega^2}, \tag{25}$$

where $\lambda = 2\pi V/\omega$ is the wave-length. The expression (25) is approximately correct if $\lambda \ll z_0$.

Thus we can write

$$h = h_0\exp(-z/z_0)\exp\{j\omega(t-z/V)\} \tag{26}$$

and

$$v = v_0\exp(-z/z_0)\exp\{j\omega(t-z/V)\}, \tag{27}$$

with

$$v_0 = \frac{Vh_0}{H_0}\left(1-j\frac{\omega c^2}{8\pi\mu\sigma V^2}\right). \tag{28}$$

The latter expression is derived from equations (5), (20), and (21).

The imaginary part indicates the phase shift of the velocity v in relation to the magnetic disturbance field h. We can also write

$$v = v_0'\exp(-z/z_0)\exp\{j\omega(t-z/V)-j\varphi\}, \tag{29}$$

where

$$v_0' = Vh_0/H_0 \tag{30}$$

and

$$\varphi = \frac{\omega c^2}{8\pi\mu\sigma V^2} = \frac{\omega c^2\rho}{2\mu^2\sigma H_0^2}. \tag{31}$$

4.4. Magneto-hydrodynamic waves as oscillations of the magnetic lines of force

A simple picture of the magneto-hydrodynamic wave motion is obtained in the following way. The magnetic lines of force are often regarded as elastic strings. The force exerted by a magnetic field may be visualized by imagining that such strings attempt to contract and, at the same time, to impose a lateral pressure upon one another. If a magnetic field exists within a body of infinite conductivity, no change in position of the lines of force can occur, since such a change would give rise to opposing currents, which would be infinitely large as a consequence of the infinite conductivity. The lines of force are thus 'frozen' in the body. As we have assumed that the fluid in question has an infinite conductivity, no relative motion between the fluid and the magnetic field is possible. At the end of § 4.31 it was shown that the motion of the fluid is the same as the motion of the lines of force. The fluid may be

considered as 'glued' to the magnetic lines of force, which may accordingly be regarded as material strings with masses equal to the mass of the fluid per line of force. Like ordinary strings, these too may begin to oscillate, and the velocity of the magneto-hydrodynamic waves may actually be derived from the equation for the oscillation of a string, which reads

$$m\frac{\partial^2 y}{\partial t^2} = S\frac{\partial^2 y}{\partial z^2} \tag{1}$$

(m = mass of string per unit length, S = tension in the string), provided that the string is parallel to the z-axis and oscillates in the direction of the y-axis (see, for example, Joos 1942, p. 162). A transverse oscillation in such a string has the wave velocity

$$V = \sqrt{(S/m)}. \tag{2}$$

If we wish to apply these equations to the oscillations of the magnetic lines of force, it is necessary to substitute for S the force with which a magnetic line of force endeavours to contract. The force per unit area at right angles to the field in a magnetic field is given by

$$P = \frac{1}{8\pi}\mu H^2. \tag{3}$$

As the number of lines of force per unit area is H, the tension per line of force will be

$$S = \frac{P}{H} = \frac{1}{8\pi}\mu H. \tag{4}$$

The mass of the string m will be the mass per line of force, i.e.

$$m = \rho/H. \tag{5}$$

If (4) and (5) are substituted in (2), however, a wave velocity is obtained which differs from 4.31 (10) by a factor $\sqrt{2}$. This is due to the fact that the magnetic lines of force also act upon one another by the above-mentioned lateral pressure. This is most simply allowed for by calculating the expressions for the energies of oscillating strings and lines of force. If the string is deformed from a straight line ($y = y_0$), so that its equation becomes

$$y = f(z), \tag{6}$$

its length between z and $z+dz$ is increased from dz to

$$dz\sqrt{\{1+(dy/dz)^2\}}.$$

Hence its total increase in length is

$$\int \left[\sqrt{\left\{1+\left(\frac{dy}{dz}\right)^2\right\}}-1\right] dz. \tag{7}$$

The tension in the string being S, this means that the potential energy of the string has increased by W, where

$$W = S\int\left[\sqrt{\left\{1+\left(\frac{dy}{dz}\right)^2\right\}}-1\right]dz, \tag{8}$$

or, if dy/dz is assumed to be $\ll 1$,

$$W = \tfrac{1}{2}S\int\left(\frac{dy}{dz}\right)^2 dz. \tag{9}$$

If a magnetic line of force is deformed from a straight line $y = y_0$ with the field strength H_0 to a form corresponding to equation (6), the consequence will be that the field strength also acquires a y-component. As

$$\frac{dy}{dz} = \frac{h_y}{H_z} \tag{10}$$

and H_z is still equal to H_0, we obtain

$$h_y = H_0\frac{dy}{dz}. \tag{11}$$

The increase in energy due to the superimposed field h_y, which is perpendicular to H_0, amounts to $\mu h_y^2/8\pi$ per unit volume and hence per unit length of a line of force

$$W = \frac{\mu}{8\pi H_0}\int h_y^2\,dz = \frac{\mu H_0}{8\pi}\int\left(\frac{dy}{dz}\right)^2 dz. \tag{12}$$

A comparison between (12) and (9) shows that we should put

$$S = \frac{1}{4\pi}\mu H_0 \tag{13}$$

instead of (4). The difference is due to the fact that in the former case we have not taken the lateral pressure of the lines of force into consideration. By inserting (13) and (5) in (2) we obtain

$$V = H_0\sqrt{(\mu/4\pi\rho)},$$

in agreement with 4.31 (10).

4.41. *Magneto-hydrodynamic waves as a special case of electromagnetic waves.* According to what has been said above, it is possible to regard magneto-hydrodynamic waves as oscillations of magnetic lines of force, 'materialized' into strings, on to which a conducting fluid has been 'glued'. This conception gives in many cases a surprisingly good survey of the phenomena. It is, however, also possible to regard the magneto-hydrodynamic waves as an extreme case of electromagnetic waves, in spite of the fact that their velocity, as shown in the example calculated in § 4.31, is often nine powers of ten lower than the velocity of light.

Consider a homogeneous magnetic field and an electromagnetic wave which travels in the direction of the field. The velocity V of the wave is given by

$$V = c(\epsilon/\mu)^{-\frac{1}{2}}, \tag{14}$$

where c is the velocity of light, ϵ the permittivity, and μ the permeability. The electric field E of the wave causes a displacement current i,

$$i = \frac{\epsilon}{4\pi}\frac{dE}{dt}. \tag{15}$$

In a vacuum we have $\epsilon = 1$. When matter is present this current is supplemented by a current in the matter. In the case of an ionized gas the current is easily found from the circuit of Fig. 3.2.

Let us first consider the case of a gas of density so high and conductivity so good that the resistances R_e and X_e in Fig. 3.2 *d* are small. The applied voltage V which corresponds to the electric field E is connected directly to the condenser, $C_M = \rho c^2/H^2$. Hence to the displacement current i_1 $(= \frac{1}{4}\pi^{-1}\,dE/dt)$ is added the current i_c $(= C_M\,dE/dt)$ which charges the condenser. The resulting current i is given by

$$i = \frac{1}{4\pi}\left(1+\frac{4\pi\rho c^2}{H^2}\right)\frac{dE}{dt}. \tag{16}$$

Comparing this with (15) we find that the current is the same as if the permittivity were

$$\epsilon = 1+4\pi\rho c^2/H^2. \tag{17}$$

Introducing this value into (14) and observing that μ still equals unity we obtain

$$V = [4\pi\rho/H^2+1/c^2]^{-\frac{1}{2}}$$

which agrees with 4.31 (10) when $4\pi\rho c^2 \gg H^2$.

The transition between electromagnetic and magneto-hydrodynamic waves can be surveyed by the help of the equivalent circuits of Fig. 3.2. When only the condenser C_M is of importance we have an undamped magneto-hydrodynamic wave. When the resistances R_e and X_M are considered also the wave is damped. An oscillation with frequency ω in a condenser C which is connected in series with a resistance R (and an inductance $L = \omega^{-2}C^{-1}$) is damped according to well-known formulae (see e.g. Harnwell, 1938, p. 421), so that its amplitude decreases to $1/e$ in a time T, where $T = 2L/R$ or

$$T = \frac{2}{\omega^2 CR}.$$

Introducing $R = R_e+X = 1/\sigma_\perp$ [see 3.22 (52)] and $C = \rho c^2/H^2$, we find the damping distance z_0, $z_0 = VT = HT(4\pi\rho)^{-\frac{1}{2}}$, or

$$z_0 = \sigma_\perp H_0^3 \pi^{-\frac{1}{2}} c^{-2} \rho^{-\frac{3}{2}} \omega^{-2}$$

in agreement with 4.32 (25).

If we decrease the density, X increases and at low densities is so large that the exchange of energy between the electronic and molecular (and ionic) components is negligible, whereas the electronic field and the electronic gas are still in contact. Then we have passed from magneto-hydrodynamic waves to 'ionospheric waves'. The equivalent circuit is that of Fig. 3.2 *b*, where the circuit elements refer to the electronic gas. When the density goes down still more, we get electromagnetic waves *in vacuo*. The circuit is reduced to the condenser C_0 which represents the displacement current.

The transition between magneto-hydrodynamic waves and electromagnetic waves has been treated mathematically by Rydbeck (1948).

4.5. Waves of arbitrary form

Up to now we have discussed some special cases in order to become familiar with the waves. We shall now enter into a more general discussion which has been given by Walén (1944).

As earlier, we neglect the displacement current. From 4.2 (2) and 4.2 (4) we eliminate E and introduce 4.2 (1). If we assume that the permeability is constant, the identity

$$\operatorname{curl}(\operatorname{curl} H) = -\Delta H + \operatorname{grad}(\operatorname{div} H)$$

reduces to $\operatorname{curl}(\operatorname{curl} H) = -\Delta H$, and we obtain

$$\operatorname{curl}[\mathbf{vH}] + \frac{c^2}{4\pi\mu\sigma}\Delta\mathbf{H} - \frac{\partial \mathbf{H}}{\partial t} = 0. \tag{1}$$

Introducing 4.2 (1) into 4.2 (5) we find

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\operatorname{grad})\mathbf{v} + \frac{\mu}{4\pi\rho}[\mathbf{H}\operatorname{curl}\mathbf{H}] = \mathbf{G} - \frac{1}{\rho}\operatorname{grad} p. \tag{2}$$

The magnetic field $$\mathbf{H} = \mathbf{H}_0 + \mathbf{h}, \tag{3}$$

and consists of the primary field $\mathbf{H}_0$, which is given and supposed to derive from currents outside the fluid, so that

$$\operatorname{curl}\mathbf{H}_0 = 0, \tag{4}$$

and the induced field $\mathbf{h}$, which is produced by the currents caused by the disturbance. Our problem is to find $\mathbf{v}$ and $\mathbf{h}$ from (1) and (2).

The general solution encounters mathematical difficulties. For the case of an *incompressible fluid with constant density* ρ *in a homogeneous magnetic field* $\mathbf{H}_0$ a solution has been given by Walén (1944). In this case we have

$$\operatorname{div}\mathbf{v} = 0, \tag{5}$$

$$\operatorname{grad} H_0 = 0. \tag{6}$$

Then the identities

$$\operatorname{curl}[\mathbf{vH}] = (\mathbf{H}\operatorname{grad})\mathbf{v}-(\mathbf{v}\operatorname{grad})\mathbf{H}-\mathbf{H}\operatorname{div}\mathbf{v}+\mathbf{v}\operatorname{div}\mathbf{H},$$
$$(\mathbf{v}\operatorname{grad})\mathbf{v} = \tfrac{1}{2}\operatorname{grad}v^2-[\mathbf{v}\operatorname{curl}\mathbf{v}],$$
$$[\mathbf{H}\operatorname{curl}\mathbf{H}] = \tfrac{1}{2}\operatorname{grad}H^2-(\mathbf{H}\operatorname{grad})\mathbf{H}$$

can be written

$$\operatorname{curl}[\mathbf{vH}] = [(\mathbf{H}_0+\mathbf{h})\operatorname{grad}]\mathbf{v}-(\mathbf{v}\operatorname{grad})\mathbf{H}, \tag{7}$$
$$(\mathbf{v}\operatorname{grad})\mathbf{v} = \operatorname{grad}\tfrac{1}{2}v^2-[\mathbf{v}\operatorname{curl}\mathbf{v}], \tag{8}$$
$$\begin{aligned}[\mathbf{H}\operatorname{curl}\mathbf{H}] &= [\mathbf{H}_0\operatorname{curl}\mathbf{h}]+[\mathbf{h}\operatorname{curl}\mathbf{h}] \\ &= \operatorname{grad}(\mathbf{H}_0\mathbf{h})-(\mathbf{H}_0\operatorname{grad})\mathbf{h}+[\mathbf{h}\operatorname{curl}\mathbf{h}].\end{aligned} \tag{9}$$

Putting
$$\mathbf{G} = -\operatorname{grad}U, \tag{10}$$
we obtain from (1)
$$(\mathbf{H}_0\operatorname{grad})\mathbf{v}-\frac{\partial\mathbf{h}}{\partial t} = -\frac{c^2}{4\pi\mu\sigma}\Delta\mathbf{h}+(\mathbf{v}\operatorname{grad})\mathbf{h}-(\mathbf{h}\operatorname{grad})\mathbf{v}, \tag{11}$$
and from (2)
$$\frac{\partial\mathbf{v}}{\partial t}+(\mathbf{v}\operatorname{grad})\mathbf{v}+\frac{\mu}{4\pi\rho}[\mathbf{H}_0\operatorname{curl}\mathbf{h}]+\frac{\mu}{4\pi\rho}[\mathbf{h}\operatorname{curl}\mathbf{h}]+\operatorname{grad}\left(U+\frac{p}{\rho}\right) = 0, \tag{12}$$
or
$$\frac{\mu}{4\pi\rho}(\mathbf{H}_0\operatorname{grad})\mathbf{h}-\frac{\partial\mathbf{v}}{\partial t}$$
$$= \frac{\mu}{4\pi\rho}[\mathbf{h}\operatorname{curl}\mathbf{h}]-[\mathbf{v}\operatorname{curl}\mathbf{v}]+\operatorname{grad}\left[U+\frac{1}{\rho}\left(p+\frac{\rho v^2}{2}\right)+\frac{\mu}{4\pi\rho}(\mathbf{H}_0\mathbf{h})\right]. \tag{13}$$

4.51. *Infinite conductivity.* If $\sigma = \infty$ an exact solution can be found for the equations (11) and (13). We put
$$\mathbf{v} = \sqrt{(\mu/4\pi\rho)}\mathbf{h}. \tag{14}$$
As
$$\tfrac{1}{2}\rho v^2+\tfrac{1}{4}\pi^{-1}\mu(\mathbf{H}_0\mathbf{h}) = \tfrac{1}{8}\pi^{-1}\mu[(\mathbf{H}_0+\mathbf{h})^2-H_0^2], \tag{15}$$
the grad term at the right side in (13) vanishes, if
$$p+\rho U+\tfrac{1}{8}\pi^{-1}\mu(\mathbf{H}_0+\mathbf{h})^2 = \text{const.}, \tag{16}$$
which means that the sum of the hydrostatic pressure p and the magnetostatic pressure $\mu H^2/8\pi$ equals $-\rho U$.

Then the equations (11) and (12) are both reduced to
$$(\mathbf{H}_0\operatorname{grad})\mathbf{h} = \left(\frac{4\pi\rho}{\mu}\right)^{\frac{1}{2}}\frac{\partial\mathbf{h}}{\partial t}. \tag{17}$$

This equation shows that *any state of motion is displaced with the velocity* $\mathbf{V}$ *in the direction of the magnetic field* $\mathbf{H}_0$, where
$$\mathbf{V} = \mathbf{H}_0\sqrt{(\mu/4\pi\rho)}. \tag{18}$$

The equations (11) and (12) are also solved by

$$\mathbf{v} = -\sqrt{(\mu/4\pi\rho)}\mathbf{h}, \tag{19}$$

the result being that the velocity becomes

$$\mathbf{V} = -\mathbf{H}_0\sqrt{(\mu/4\pi\rho)}. \tag{20}$$

Hence, if the mass velocity v is antiparallel to the magnetic disturbance field h, the wave velocity V is also antiparallel to the given magnetic field H_0.

In the calculations no second-order terms have been neglected. Consequently the result holds even if $h > H_0$.

4.6. Magneto-hydrodynamic whirl rings

If a ring in the fluid is set in motion with the velocity $2v$, it gives rise to magneto-hydrodynamic waves. As Walén has shown, the result is that the ring is split into two equal rings, each having the hydrodynamic velocity v. One of them migrates with the velocity $+V$ (in the direction of the given magnetic field H_0), the other is displaced with the velocity $-V$ (antiparallel to H_0).

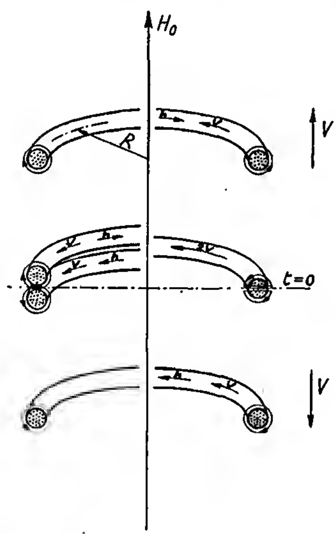

FIG. 4.6. An initial hydrodynamic wave is split into two magneto-hydrodynamic rings, one migrating in the direction $+\mathbf{H}_0$, the other in the direction $-\mathbf{H}_0$. (After Walén.)

According to 4.51 (14) we have

$$\frac{\rho v^2}{2} = \frac{h^2\mu}{8\pi}, \tag{1}$$

which means that the kinetic energy of the ring equals its magnetostatic energy. As a consequence of this the centrifugal force due to the curvature of the ring is always compensated by the magnetostatic pull of the lines of force.

Another consequence is that the hydrodynamic reduction in pressure, which equals $\frac{1}{2}\rho v^2$, is always compensated by the magnetostatic pressure $\frac{1}{8}\mu h^2/\pi$.

The hydrodynamic motion v, with the component $v_\perp$ perpendicular to H_0, produces an electric polarization $\mathbf{E}'$ within the ring.

$$\mathbf{E}' = (\mu/c)[\mathbf{v}\mathbf{H}_0]. \tag{2}$$

Consider a plane parallel to H_0 and E (Fig. 4.7). If we integrate along

a closed fixed curve $ABCA$, an electromotive force

$$V = \oint E\,ds = -\frac{1}{c}\frac{\partial\phi}{\partial t} \tag{3}$$

is produced due to the change in magnetic flux ϕ when the ring proceeds. During the time Δt the ring is displaced $V\Delta t$, so that $\Delta\phi$, the increase in flux through the surface $ABCA$, is equal to $\mu h_\perp bV\Delta t$, where $h_\perp$ is the component of h perpendicular to H_0 and b is the breadth of the ring (distance AC). As $h_\perp = v_\perp H_0/V$, we have

$$V = \oint E\,ds = -\frac{1}{c}\mu h_\perp bV = -\frac{\mu}{c}v_\perp H_0 b, \tag{4}$$

which means that the transformer induced voltage exactly compensates the voltage due to polarization. As this holds for any curve ABC, no electric current is produced in the surrounding medium. Hence the induced e.m.f. could be supposed to be due to an induced electric field E inside the ring,

$$\mathbf{E} = -\mu c^{-1}[\mathbf{vH}_0], \tag{5}$$

which exactly neutralizes the polarization $\mathbf{E}'$.

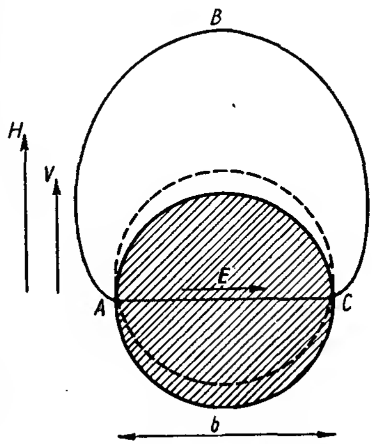

FIG. 4.7. Showing the cross-section of a branch of a magneto-hydrodynamic ring. The hydrodynamic flow perpendicular to the paper causes a motional induced e.m.f. between A and C, which in the ideal case is exactly compensated by a transformer induced e.m.f. due to the change of flux in the fixed circuit ABC.

What has been said above is exact if the conductivity is infinite. If so, the shape of the rings remains unchanged during their displacement. If the conductivity is finite, but so large that the damping is not very important, the above considerations are approximate. The main effect of the finite conductivity is that the cross-section of the rings increases. This follows the same law as the penetration of a magnetic field into a conductor, a phenomenon well known from the 'skin-effect'. The velocity of the ring 'diffuses' into the surroundings. The limitation of the ring becomes more and more diffuse. As an approximate measure of the diffusion we could say that the cross-section S of the magneto-hydrodynamic ring increases at the rate

$$\frac{dS}{dt} = \frac{c^2}{2\pi\sigma\mu}. \tag{6}$$

The initial cross-section S is consequently doubled when the ring has travelled a distance

$$z_1 = Vt = \frac{2\pi\sigma\mu V}{c^2} S. \tag{7}$$

This distance must be related to the damping distance z_0 of 4.32 (25). If the ring is regarded as a Fourier series of sine waves the most important wave-lengths are given by $(\lambda/\pi)^2 \approx S$. Comparing (7) and 4.32 (25) we find that $z_1 \approx z_0$ as expected.

4.7. Waves in inhomogeneous magnetic field, compressible fluid with variable density

An exact treatment of the general case when the magnetic field is inhomogeneous and the density a function of the coordinates encounters mathematical difficulties. If the relative variations of the field and of the density over the space occupied by a magneto-hydrodynamic ring are small, we can expect our equations to be approximately valid. Thus each element of the ring travels with velocity V equal to $H_0\sqrt{(\mu/4\pi\rho)}$, which is variable. Consider a cylindrical element having the base surface ΔF perpendicular to H_0 and the height Δz parallel to H_0. When the magnetic field changes, the flux $H_0\Delta F$ through the base surface will remain constant, which means that ΔF varies inversely as H_0. Further, the height Δz varies as V when H_0 and ρ change. Thus the volume $\Delta\tau = \Delta F \Delta z$ is independent of H_0 and proportional to $\rho^{-\frac{1}{2}}$. Further, the kinetic energy $\frac{1}{2}\rho v^2 \Delta\tau$ and the magnetic energy $\frac{1}{8}\pi^{-1}\mu h^2 \Delta\tau$ should be conserved. Thus we find the following variations when ρ and H_0 change:

$$\text{Hydrodynamic velocity } v = \text{const.}\, \rho^{-\frac{1}{4}}, \tag{1}$$

$$\text{Induced magnetic field } h = \text{const.}\, \rho^{\frac{1}{4}}, \tag{2}$$

Cross-section of a ring projected on a plane

$$\text{perpendicular to the field } F = \text{const.}\, H^{-1}. \tag{3}$$

The shape of a ring may change because ΔF and Δz vary in different ways. It must be observed that even the surface ΔF may change its shape if H_0 is inhomogeneous (which means that if, for example, the projection of a ring perpendicular to the magnetic field is circular at one instant, it does not necessarily remain so).

The effect of a variation of density in the presence of a gravitational field will be discussed in the following section.

When the fluid is compressible the change in hydrostatic pressure may be of importance. In a magneto-hydrodynamic ring the pressure

is $\mu h^2/8\pi$ less than in the environment. When this pressure is small in comparison to the total pressure no important new effect will occur.

A compressible liquid is able to propagate sound waves. These are longitudinal (material velocity and wave velocity parallel), in contrast to magneto-hydrodynamic waves, which like electromagnetic waves are transverse (material velocity perpendicular to wave velocity).† In a conducting medium the presence of a magnetic field may change the velocity of sound and also introduce a damping.

4.8. Gravitational effect

If the density is variable a gravitational field affects a magneto-hydrodynamic ring. Let us suppose that the gravitation g acts in the direction of the negative ζ-axis of an orthogonal coordinate system (ξ, η, ζ). Consider an incompressible liquid, the density $\rho(\zeta)$ of which is a function of ζ alone (at a certain time). Suppose that a volume element $d\Omega$ initially situated at ζ_0, and thus having the density $\rho(\zeta_0)$, is displaced to ζ_1. When it is situated at ζ, where the density of the liquid around it is $\rho(\zeta)$, the force acting upon it is $g\,d\Omega[\rho(\zeta)-\rho(\zeta_0)]$. Thus the energy gained through the displacement is

$$dW = g\,d\Omega \int\limits_{\zeta_0}^{\zeta_1} [\rho(\zeta)-\rho(\zeta_0)]\,d\zeta. \tag{1}$$

Let us suppose that in the region under consideration the density can be expressed as a linear function of ζ:

$$\rho(\zeta) = \rho_0+(2\Theta/g)(\zeta-\zeta_0), \tag{2}$$

ρ_0 and Θ being constants. Then we have

$$dW = \Theta(\zeta_1-\zeta_0)^2 d\Omega. \tag{3}$$

We shall now discuss how this affects a whirl ring proceeding with the wave-velocity V in the direction of a homogeneous magnetic field H_0, which is parallel to the z-axis of an orthogonal coordinate system (x, y, z). The angle between the ζ- and the z-axes is denoted by α. An exact treatment of the motion of a whirl ring when the density is variable encounters great analytical difficulties, but if the relative change in density over the diameter of the ring is small—as we shall assume—an approximate treatment seems permissible.

The magneto-hydrodynamic ring is characterized by a material

† The material velocity of a magneto-hydrodynamic wave may have a component parallel to the magnetic field and hence parallel to the wave velocity, but this produces no force.

velocity v which is a function of the coordinates x, y, and $z-Vt$ (where V is the wave velocity). Thus the *state of motion* is displaced with the velocity V in the direction of the magnetic field H_0, but the matter within the ring moves with the velocity $\mathbf{v}$, which in general is different from the wave-velocity $\mathbf{V}$.

Consider a small volume element $d\Omega$ which during the interval t to $t+\tau$ forms a part of the whirl ring. Before t it is at rest at a point P_0 whose radius vector is $\mathbf{r}_0$. During the interval τ it moves with the velocity $\mathbf{v}$, which may vary during the time τ, and is hence displaced the distance

$$\mathbf{s} = \int_t^{t+\tau} \mathbf{v}\, dt. \tag{4}$$

After $t+\tau$ it is again at rest but now at a point P_1, given by

$$\mathbf{r}_1 = \mathbf{r}_0+\mathbf{s}. \tag{5}$$

Denoting the ζ-components of $\mathbf{r}_0$ and $\mathbf{r}_1$ by ζ_0 and ζ_1, we find the work gained by the whirl in displacing the volume $d\Omega$ from formula (3).

Let us denote the projection upon the xy-plane of a surface element of the forward side of the whirl ring by dS. During the time interval between t and $t+dt$ the ring proceeds a distance $V\,dt$ and consequently takes in a volume $\iiint d\Omega$, where

$$d\Omega = V\, dSdt. \tag{6}$$

Introducing (6) into (3) and denoting the ζ-component of $\mathbf{s}$ by $\Delta\zeta$, so that

$$\Delta\zeta = \zeta_1-\zeta_0, \tag{7}$$

we find after integration the rate of energy gain dW/dt of the whole whirl:

$$\frac{dW}{dt} = \Theta V \iint (\Delta\zeta)^2\, dS. \tag{8}$$

We shall discuss some special cases.

4.81. Suppose that the material velocity $\mathbf{v}$ of every given element $d\Omega$ is constant during the interval τ, and denote by v_ζ the projection of $\mathbf{v}$ upon the ζ-axis. The velocity $\mathbf{v}$ may be (and is in general) a function of the coordinates of the surface element dS through which it is taken in. A line parallel to the z-axis (parallel to H_0) will be intersected by the boundary surface of the ring at two points whose distance apart we call Δz. Suppose further that $|s| \ll \Delta z$. Then we have

$$\Delta z = V\tau, \tag{9}$$

$$\Delta\zeta = v_\zeta \tau = v_\zeta V^{-1}\Delta z, \tag{10}$$

which gives by (8)
$$\frac{dW}{dt} = \frac{\Theta}{V}\iint v_\zeta^2(\Delta z)^2\,dS, \tag{11}$$

where v_ζ and Δz are functions of x and y.

4.82. Suppose that the whirl is circular and limited by two planes parallel to the xy-plane at an intermediate distance $c = V\tau$ and two coaxial cylindrical surfaces r_1 and r_2 parallel to the z-axis (see Fig. 4.8). We assume that $b = r_2 - r_1 \ll r_2$ and denote the mean radius by r.

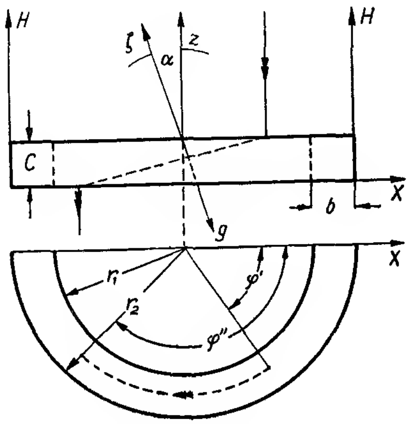

FIG. 4.8. Whirl perpendicular to the magnetic field.

Orienting the x-axis so that the ζ-axis falls in the xz-plane and denoting the angle between r and x by φ, we have:

$$dS = br\,d\varphi, \tag{12}$$

$$\Delta\zeta = -\Delta x \sin\alpha = r\sin\alpha(\cos\varphi' - \cos\varphi'') = 2r\sin\alpha\sin\varphi\sin\tfrac{1}{2}\Delta\varphi, \tag{13}$$

where φ' and φ'' are the angles at which a volume element enters and leaves the ring, and $\varphi = \frac{1}{2}(\varphi'+\varphi'')$; $\Delta\varphi = \varphi''-\varphi'$. Thus we have, according to (8),

$$\frac{dW}{dt} = 4\Theta Vbr^3\sin^2\alpha\sin^2(\tfrac{1}{2}\Delta\phi)\int_0^{2\pi}\sin^2\varphi\,d\varphi = 4\pi\Theta Vbr^3\sin^2\alpha\sin^2(\tfrac{1}{2}\Delta\varphi). \tag{14}$$

As
$$\Delta\varphi = \frac{v\tau}{r} = \frac{c}{r}\frac{v}{V} \tag{15}$$

and
$$W = \tfrac{1}{2}mv^2 = 2\pi rbc\rho\tfrac{1}{2}v^2, \tag{16}$$

we have, for $\Delta\varphi \ll 1$,
$$\frac{dv}{v} = \frac{\Theta c}{2\rho V}\sin^2\alpha\,dt, \tag{17}$$

indicating that v increases exponentially for small values of $\Delta\varphi$. For larger values the increase is less rapid, and when

$$\Delta\varphi = 2n\pi \quad (n = 1, 2, 3,...)$$

the increase in kinetic energy of the ring becomes zero according to (14). For these values the matter in the ring rotates exactly one turn and is left after the passage of the whirl at exactly the same point where it was before the whirl arrived. Consequently, *if a whirl perpendicular to the magnetic field starts with a small velocity v, it increases until $\Delta\varphi = 2\pi$ (which means $v = 2\pi Vr/c$) but not more.*

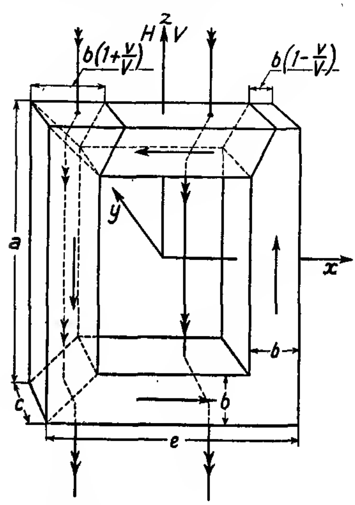

Fig. 4.9. Whirl parallel to the magnetic field.

4.83. We shall now treat the case when the plane of the whirl is parallel to the magnetic field. In order to avoid somewhat complicated geometrical factors, let us assume that in a plane (xz-plane) parallel to the field the whirl has the rectangular shape shown in Fig. 4.9. The breadth b is supposed to be small in comparison with a and e. Perpendicular to this plane it has the thickness c.

All the matter taken in by the whirl through the front surface ec immediately acquires the velocity whose components are given by

$$v_x = -v, \quad v_z = 0. \tag{18}$$

At the same time the whirl proceeds in the z-direction with the velocity V. The result of this is that matter taken in at a distance smaller than $b(1+v/V)$ from the left edge (see Fig. 4.9) of the front passes through the whole left branch of the whirl and is left at the rear at a point symmetrical to that at which it entered the whirl. Thus the displacement in the x-direction is zero, but the z-displacement is

$$\Delta_1 z = -v\tau = -v\{a/(V+v)\}. \tag{19}$$

If $V > v$ there is similarly a part $b(1-v/V)$ near the right edge of the front surface from which matter passes through the right branch to the rear, the displacement in the z-direction being given by

$$\Delta_2 z = v\tau = v\{a/(V-v)\} \tag{20}$$

and in the x-direction by $\Delta x = 0$. Matter entering on the line $a-2b$

between these two surfaces is displaced in the x-direction only

$$\Delta_3 x = -v\tau = -v(b/V). \tag{21}$$

In the same way matter entering at the inner side of the back branch is displaced $\Delta_4 x$

$$\Delta_4 x = v(b/V). \tag{22}$$

Let us at first suppose that the x-component of gravitation is zero. Putting

$$\Theta_z = \Theta \sin^2\alpha, \tag{23}$$

we have from (8)

$$dW/dt = \Theta_z V[(\Delta_1 z)^2 S_1 + (\Delta_2 z)^2 S_2], \tag{24}$$

where

$$S_1 = cb(1+v/V); \qquad S_2 = cb(1-v/V), \tag{25}$$

which gives

$$dW/dt = 2\Theta_z Va^2bc\, v^2/(V^2-v^2). \tag{26}$$

If the x-component of gravitation differs from zero, there should formally be an additive increase in energy

$$(dW/dt)_x = \Theta_x V\{(\Delta_3 x)^2 + (\Delta_4 x)^2\} S_3, \tag{27}$$

where

$$S_3 = c(a-2b). \tag{28}$$

The matter displaced $\Delta_3 x$ by the forward branch, however, is displaced back again $\Delta_4 x$ ($= -\Delta_3 x$) by the rear branch, so that the resulting displacement is zero. The expression (27) is correct if the matter in question is mixed with adjacent matter when at rest between the two displacements in such a way that, when the rear branch reaches it, it has the density which is normal at its new level. On the other hand, if no mixing takes place, the resultant work to displace and displace back again must be zero. As in our treatment we have not taken account of any diffusion and mixing between the matter in the whirl and that in the environment, it is consistent to neglect the above mixing also and put $(dW/dt)_x$ equal to zero.

If the whirl is square—which represents our approximation to the normal circular whirl—we have $a = e$ and

$$W = \tfrac{1}{2}mv^2 = 2abc\rho v^2. \tag{29}$$

This gives with (26)

$$\frac{dv}{dt} = K\frac{v}{1-v^2/V^2}, \tag{30}$$

where

$$K = \Theta_z a/2\rho V. \tag{31}$$

As long as $v \ll V$ the increase (or decrease) in v is almost exponential:

$$v = v_0 e^{Kt}. \tag{32}$$

When v approaches V the increase becomes more rapid and is infinite for $v = V$. This depends upon the fact that the motion $V-v$ in relation to the whirl in the right branch (see Fig. 4.9) becomes zero, so that matter is 'captured' in this branch and displaced an infinite distance

($\Delta z = \infty$). If $v > V_1$ the motion in relation to the whirl goes upward in the right branch and its matter is left at the inside of the forward branch, being again taken in at the inside of the rear branch and brought back to the right branch. In this case also matter is 'captured' by the whirl, so that Δz becomes infinite. In reality the diffusion and mixing of the matter in the whirl and in its environment, of which we have taken no account, makes the increase finite although very large.

A tentative estimation of the rate of increase in this case will be made here. Suppose that the matter captured by the whirl becomes mixed with the matter in the environment after a certain time τ. Then the whirl transports the 'captured' matter a distance Δz,

$$\Delta z = V\tau. \tag{33}$$

At a certain instant the density of the left branch (supposing the same direction of the whirl as above) of a whirl situated at $z = z$ is approximately $\rho(z)$, whereas the density of the right branch is $\rho(z-\Delta z)$. The force accelerating the whirl is $bcag\{\rho(z)-\rho(z-\Delta z)\}$, and the mass which is accelerated equals $4\rho abc$. In consequence of (2) and (33) we obtain

$$dv/dt = (\Theta_z/2\rho)V\tau. \tag{34}$$

As Δz ($= V\tau$) is supposed to be much larger than a, this increase is much swifter than that given by (30) except when v is very close to V. Thus when $v > V$ the acceleration is linear but in general much larger than when $v < V$.

When the plane of the whirl makes an angle β with the magnetic lines of force, the matter in the whirl has a velocity component $V\sin\beta$ perpendicular to the plane of whirl. When the thickness of the whirl is c, the maximum value of the time τ during which matter can remain captured by the whirl is given by

$$\tau < c/(V\sin\beta).$$

If for $\beta = 0$ the value of Δz is, for example, $5a$ and $c = a/5$, the above inequality shows that τ diminishes if β is increased above

$$\beta = \sin^{-1}(1/25) = 2^\circ.$$

Consequently, a whirl which makes a larger angle with the magnetic lines of force is accelerated at a lower rate.

This shows that the acceleration process which we have studied is especially favourable for whirls making only small angles with the magnetic field.

REFERENCES

ALFVÉN, H., 1942: 'On the Existence of Electromagnetic-hydrodynamic Waves', *Ark. f. mat., astr. o. fysik*, **29 B**, No. 2; *Nature*, **150**, 405.

HARNWELL, G. P., 1938: *Principles of Electricity and Electromagnetism*, New York.

JOOS, G., 1942: *Lehrbuch der theoretischen Physik*, Leipzig. (English edition: *Theoretical Physics*, London and Glasgow.)

RYDBECK, O., 1948: 'On the Propagation of Waves in Inhomogeneous Media', *Trans. Chalmer's University of Technology*, Gothenburg, Sweden, **74**.

WALÉN, C., 1944: 'On the Theory of Sunspots', *Ark. f. mat., astr. o. fysik*, **30 A**, No. 15; **31 B**, No. 3.

Note added in proof: Magneto-hydrodynamic waves have recently been demonstrated by S. Lundquist (*Nature*, 1949). He excites the waves mechanically in mercury in a magnetic field of the order 10,000 gauss.

V

SOLAR PHYSICS

5.1. To give a short review of well-known facts, the sun is a sphere consisting of hot ionized non-degenerate gas, mainly hydrogen. Its energy is produced through nuclear processes in a rather small region near the centre where the temperature is about $20 . 10^6$ degrees and the density about 100 times that of water. The energy is transferred outwards through radiation, in some regions also, in part, through convection.

The solar density decreases outwards continuously. In the outer parts of the sun three different layers can be discerned: the photosphere, the chromosphere, and the corona. The photosphere is the layer from which the visual solar radiation is directly (without further re-emission) transmitted. It is a thin layer (only some 10^8 cm. thick) situated at a central distance of $6{\cdot}95 . 10^{10}$ cm., which figure is called the 'solar radius'. It should be remarked that if the sun is observed in other wave-lengths than those of ordinary light its apparent radius may be different. For example, for radio waves it is much bigger, because these are absorbed high up in the corona.

In the photosphere the density falls very rapidly so that the solar limb is very sharp. The radiation from inside the limb is a continuous spectrum with the Fraunhofer absorption lines.

Above the photosphere comes the chromosphere. When seen at the limb (immediately outside the border) it gives a spectrum of emission lines ('flash spectrum') the intensity of which is much lower than that of the photosphere. The chromosphere has an irregular, rapidly changing, structure and is by some authors considered as consisting of a multitude of small jets or prominences. The height of the chromosphere is about 10^9 cm.

The corona is situated outside the chromosphere. It has a continuous spectrum due to scattered photospheric light and, superimposed upon it, some emission lines. These derive from very highly ionized atoms. The high ionization indicates that the coronal temperature is about 10^6 degrees. The corona can be traced out to about 10 times the solar radius.

The layers below the photosphere are not accessible for direct observation. What we know of them is derived theoretically. Up to the discovery of the magneto-hydrodynamic waves the total mass, radius, and energy flow were the only observational data which a theory of the sun's interior must satisfy.

Theories of solar structure start from the assumption that most energy is produced very close to the solar centre (point-source model). The concentration of the energy production is due to the fact that the nuclear reactions delivering the energy are extremely sensitive to the temperature T (energy production proportional to T^{16} or T^{18} under the conditions in question), so that energy is released only in a very narrow region where the sun is hottest.

From the central energy source the energy is transmitted outwards to the photosphere which emits it into space. The energy transmission within the sun is mainly carried out by radiation, the energy being emitted, absorbed, and re-emitted by subsequent solar layers. In the central part, however, radiation cannot alone transmit all the energy (the radiation depending on temperature as the fourth power only, whereas the energy production goes with a much higher power) so that convection must set in. Convection takes place also in the photosphere as shown by observation (granulation). This is due to an instability caused by the ionization of hydrogen.

The simplest theory of the sun's interior leads to the 'polytropic model'. Refined theories taking account of the variation of opacity with temperature have lately been developed. The assumptions concerning the chemical composition differ. Blanch, Lowan, Marshak, and Bethe (1941) assume the sun to contain 35 per cent. hydrogen and no helium, the rest being 'Russell mixture' of heavier elements. Their model is given in Table 5.1 and Fig. 5.1. According to Schwarzschild (1946) the sun contains 47 per cent. hydrogen, 41 per cent. helium, and 12 per cent. Russell mixture. In Schwarzschild's model the density in the inner parts is slightly higher, in the outer parts considerably lower than in Blanch–Lowan–Marshak–Bethe's model.

Wildt's discovery (1939) of the importance of the negative hydrogen ion has recently led to a revision of ideas of the chemical constitution of the outer layers. According to B. Strömgren (1940) the photosphere and layer outside it consist almost entirely of hydrogen. The ratio of metals to hydrogen is only 1:8,000. The content of helium, carbon, oxygen, and nitrogen is rather small. When calculating the pressure, conductivity, etc., of the outer layers it should be permissible to assume that they consist of pure hydrogen. Whether there really is a difference in chemical constitution between the inner and outer layers depends upon the degree of convection in the sun. If this should turn out to be great even in the intermediate layers, no considerable difference is possible (compare § 5.25).

TABLE 5.1

Central distance R 10^{10} cm.	Temperature T 10^6 °K.	Density ρ $g.\,cm.^{-3}$	Scale-height ζ_1 10^{10} cm.	Magnetic field H_a at axis gauss	Velocity V_a of magneto-hydrodynamic waves at axis cm. sec.$^{-1}$	Lowest transmitted frequency $\omega_c^{(a)}$ 10^{-10} sec.$^{-1}$	Conductivity σ/c^2 10^{-5} e.m.u.	$A = \omega^2 z_0$, z_0 = damping distance 10^3 cm. sec.$^{-2}$	Highest transmitted frequency $\omega = (A/\zeta_1)^{\frac{1}{2}}$ 10^{-5} sec.$^{-1}$
0	25·7	110	∞	2000	53·7	0	64	2·56	0
1·0	19·2	69·7	1·20	2000	67·6	28·2	41·3	3·200	51·8
1·5	14·7	36·7	0·65	2000	93·1	71·7	27·7	5·630	93·1
2·0	11·1	15·3	0·54	1049	75·6	70·0	18·2	1·990	60·8
2·5	8·60	5·90	0·52	536	62·2	59·9	12·4	0·751	38·0
3·0	6·72	2·23	0·52	312	59·0	56·8	8·52	0·442	29·2
3·5	5·18	0·85	0·53	194·1	59·4	56·0	5·75	0·304	24·1
4·0	3·93	0·34	0·53	131·1	63·4	59·8	3·82	0·245	21·6
4·5	2·85	0·132	0·50	92·1	71·5	71·5	2·36	0·218	20·9
5·0	2·05	0·0435	0·44	67·14	90·8	103·3	1·44	0·272	24·9
5·5	1·40	0·0126	0·35	50·44	126·8	181·0	0·812	0·416	34·5
6·0	0·85	0·0021	0·24	38·85	239	498	0·384	1·320	74·2
6·5	0·37	0·00014	0·17	30·56	727	2140	0·11	10·7	252
6·7	0·20	0·0000395	0·15	27·90	1252	4170	0·044	21·8	381
6·8	0·118	0·000018	0·13	26·68	1775	6820	0·017	23·9	429

Remarks on Table 5.1

1. Solar radius $R_0 = 6{\cdot}96.10^{10}$ cm.
2. Temperature from Blanch, Lowan, Marshak, and Bethe.
3. Density from Blanch, Lowan, Marshak, and Bethe.
4. Scale-height computed from their data.
5. Magnetic field at magnetic axis H_a, according to homogeneous-field-dipole-field model (5.22). Dipole moment $a = 4{\cdot}2.10^{33}$ gauss cm.^{3}, central field 2,000 gauss.
6. Velocity V_a of magneto-hydrodynamic waves at magnetic axis. $V_a = H_a\,(4\pi\rho)^{-\frac{1}{2}}$.
7. Lowest magneto-hydrodynamic frequency ω_c which is not reflected due to the change in refractive index (see 5.23). $\omega_c = V_a/2\zeta_1$.
8. Conductivity from Cowling. σ in e.s.u., hence σ/c^2 in e.m.u. Compare conductivity of copper: $\sigma/c^2 = 58.10^{-5}$ e.m.u.
9. Due to the finite resistance the amplitude of a wave with frequency ω sec.$^{-1}$ is damped to $1/e$ after having travelled z_0 cm.
10. Highest frequency which passes the layer without considerable damping.

Note that Blanch, Lowan, Marshak, and Bethe have assumed the sun to contain 35 per cent. H (by weight) and Russell mixture.

The sun rotates around its axis, but the rotation is non-uniform, being more rapid near the equator than at high latitudes.

5.2. Electromagnetic properties of the sun

From an electromagnetic point of view the sun can be considered as a good electrical conductor at least up to the photosphere. Even in the

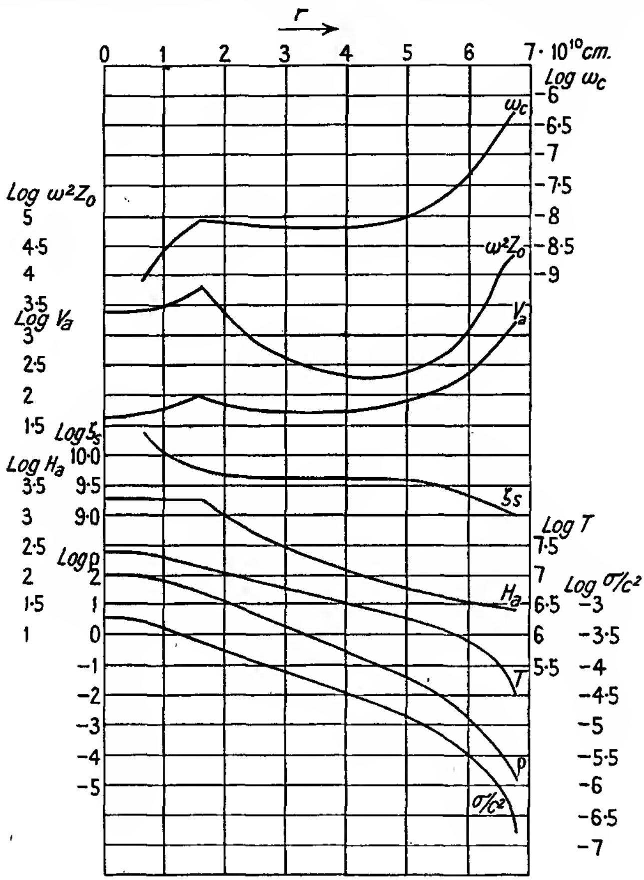

FIG. 5.1. Solar models.

chromosphere and corona the conductivity parallel to the magnetic field is high, but the cross-conductivity decreases with decreasing density. The conductivity will be discussed in § 5.21.

In all electromagnetic phenomena the general magnetic field of the sun is of essential importance. The arguments for the existence of a general field and its probable properties are reviewed in § 5.22.

TABLE 5.2.

Central distance R 10^{10} cm.	Temperature T 10^6 °K.	Density N Particles/cm.^{3}	Scale-height ζ_1 10^{10} cm.	Magnetic field H_a at axis gauss	Velocity V_a of magneto-hydrodynamic waves at axis cm. sec.$^{-1}$	Lowest transmitted frequency $\omega_c^{(a)}$ cm. sec.$^{-1}$	Conductivity σ/c^2 e.m.u. $\sigma_\parallel/c^2$	$\sigma_\perp/c^2$	$A = \omega^2 z_0$, z_0 = damping distance cm. sec.$^{-2}$	Highest transmitted frequency $\omega = (A/\zeta_1)^{\frac{1}{2}}$ sec.$^{-1}$
6·45	0·415	10^{20}	0·17	31·25	681	$2{\cdot}01.10^{-7}$	$1{\cdot}31.10^{-6}$	$1{\cdot}31.10^{-6}$	$1{\cdot}04.10^{4}$	$2{\cdot}47.10^{-3}$
6·82	0·090	10^{19}	0·11	26·45	1820	$8{\cdot}27.10^{-7}$	$1{\cdot}43.10^{-7}$	$1{\cdot}43.10^{-7}$	$2{\cdot}17.10^{4}$	$4{\cdot}45.10^{-3}$
6·944	0·0116	10^{18}	0·0087	25·08	5500	$3{\cdot}16.10^{-5}$	$2{\cdot}14.10^{-8}$	$1{\cdot}99.10^{-8}$	$8{\cdot}32.10^{4}$	$3{\cdot}09.10^{-2}$
6·951	0·00552	10^{17}	0·0022	25·00	$17{\cdot}2.10^{3}$	$3{\cdot}91.10^{-4}$	$6{\cdot}51.10^{-9}$	$3{\cdot}88.10^{-9}$	$4{\cdot}96.10^{5}$	0·149
6·954	0·00486	10^{16}	0·00050	24·96	$54{\cdot}9.10^{3}$	$5{\cdot}49.10^{-3}$	$5{\cdot}37.10^{-9}$	$8{\cdot}15.10^{-10}$	$3{\cdot}38.10^{6}$	0·822
6·971	0·035	10^{15}	0·011	24·77	$17{\cdot}0.10^{4}$	$7{\cdot}90.10^{-4}$	$1{\cdot}04.10^{-7}$	$1{\cdot}02.10^{-7}$	$1{\cdot}25.10^{10}$	10·7
6·996	0·035	10^{14}	0·011	24·54	$53{\cdot}8.10^{4}$	$2{\cdot}47.10^{-3}$	$1{\cdot}04.10^{-7}$	$2{\cdot}92.10^{-8}$	$1{\cdot}14.10^{11}$	32·2
7·021	0·035	10^{13}	0·011	24·25	$16{\cdot}8.10^{5}$	$7{\cdot}71.10^{-3}$	$1{\cdot}04.10^{-7}$	$4{\cdot}10.10^{-10}$	$4{\cdot}88.10^{10}$	20·9
7·050	0·035	10^{12}	0·011	23·95	$52{\cdot}5.10^{5}$	$2{\cdot}41.10^{-2}$	$1{\cdot}04.10^{-7}$	$4{\cdot}26.10^{-12}$	$1{\cdot}55.10^{10}$	11·9
7·071	0·035	10^{11}	0·011	23·74	$16{\cdot}6.10^{6}$	$7{\cdot}62.10^{-2}$	$1{\cdot}04.10^{-7}$	$4{\cdot}31.10^{-14}$	$4{\cdot}95.10^{9}$	6·71
7·096	0·035	10^{10}	0·011	23·48	$51{\cdot}5.10^{6}$	$2{\cdot}36.10^{-1}$	$1{\cdot}04.10^{-7}$	$4{\cdot}41.10^{-16}$	$1{\cdot}52.10^{9}$	3·71
7·121	0·035	10^{9}	0·011	23·25	$16{\cdot}0.10^{7}$	$7{\cdot}34.10^{-1}$	$1{\cdot}04.10^{-7}$	$4{\cdot}49.10^{-18}$	$4{\cdot}62.10^{8}$	2·05
7·87	1·17	10^{8}	0·772	17·15	$37{\cdot}6.10^{7}$	$4{\cdot}86.10^{-2}$	$2{\cdot}01.10^{-5}$	$4{\cdot}65.10^{-22}$	$6{\cdot}2.10^{5}$	$8{\cdot}94.10^{-3}$
10·6	1·09	10^{7}	1·67	7·02	$48{\cdot}3.10^{7}$	$2{\cdot}90.10^{-2}$	$1{\cdot}81.10^{-5}$	$3{\cdot}08.10^{-23}$	$7{\cdot}5.10^{4}$	$2{\cdot}12.10^{-3}$
15·5	0·745	10^{6}	2·57	2·25	$49{\cdot}4.10^{7}$	$1{\cdot}94.10^{-2}$	$1{\cdot}03.10^{-5}$	$5{\cdot}30.10^{-24}$	$1{\cdot}6.10^{4}$	$7{\cdot}89.10^{-4}$
22·7	0·509	10^{5}	3·79	0·719	$49{\cdot}4.10^{7}$	$1{\cdot}30.10^{-2}$	$5{\cdot}75.10^{-6}$	$9{\cdot}09.10^{-25}$	$2{\cdot}75.10^{3}$	$2{\cdot}69.10^{-4}$
33·4	0·345	10^{4}	5·58	0·225	$49{\cdot}4.10^{7}$	$8{\cdot}85.10^{-3}$	$3{\cdot}22.10^{-6}$	$1{\cdot}68.10^{-25}$	$5{\cdot}09.10^{2}$	$0{\cdot}96.10^{-4}$

Remarks on Table 5.2

Compare Table 5.1. Chemical composition: according to Strömgren mostly hydrogen.

Photosphere: Density and temperature from Waldmeier (1945), interpolated to Blanch, Lowan, Marshak, and Bethe values. Conductivity from § 3.23.

Chromosphere: Scale-height is taken from Wildt (1947) and assumed to be constant. Temperature is computed from scale-height assuming hydrostatic equilibrium (see § 5.4.) Central distance computed from scale-height.

Corona: Density as function of radius from Baumbach's formula, revised by van der Hulst. Temperature and scale-height computed from this formula, assuming hydrostatic equilibrium.

As the sun consists of a conducting medium with a magnetic field, the conditions for magneto-hydrodynamic waves are favourable (5.23). Except in the highest layer (especially the corona) most electromagnetic phenomena must be of magneto-hydrodynamic character. Electro-

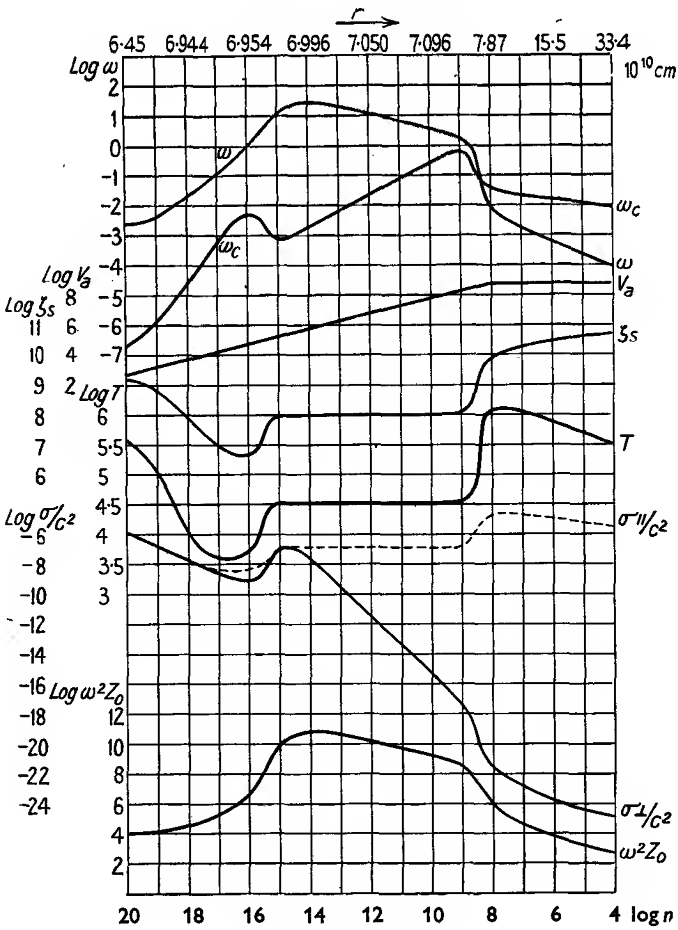

FIG. 5.2. Solar models.

magnetic effects occur as a product of relative motions and in general tend to brake the motions. Hence the non-uniform rotation of the sun can persist as a stationary state only if certain conditions are satisfied (5.24).

After a general survey of those solar phenomena which are likely to be of electromagnetic origin (§ 5.25), some are discussed in more detail in the following sections (5.3–5.9).

5.21. *The conductivity.* The conductivity in the sun is given by Cowling (1945 *b*) (compare equation 3.23 (66)) as

$$\sigma/c^2 = 1{\cdot}35.10^{-14}\,T^{\frac{3}{2}}/Z. \tag{1}$$

He puts 2·3 for the value of the mean ionization Z and uses the temperature found by Blanch, Lowan, Marshak, and Bethe. The values he obtains are given in Table 5.1 and Fig. 5.1. For the outer layer the same formula may be used at least for an estimate of the order of magnitude. As these layers are supposed to consist almost entirely of hydrogen, we put $Z = 1$. It must be observed that in the photosphere the gas is not completely ionized (as is the case in deeper layers as well as in the corona). Because ions have larger collisional cross-sections than neutral atoms, the latter do not affect the conductivity unless there are more than about 10^3 neutral atoms per ion (see § 3.23). In the photosphere the metals are ionized, but not hydrogen. The abundance of metals is estimated as 1:8,000 of that of hydrogen (see B. Strömgren), and this figure also represents the ratio of ions to atoms. Hence the atoms may diminish the conductivity but the decrease is less than one power of 10. As this effect is somewhat uncertain, Table 5.1 contains values computed directly from Cowling's formula (1).

In the photosphere and below it the value of $\omega_e \tau_e$ is very small so that the cross-conductivity equals the parallel conductivity.† In the chromosphere and corona it becomes considerable or large, so that we must distinguish between parallel conductivity and cross-conductivity. According to equation 3.23 (65) we have

$$\omega_e \tau_e = 10^6 \frac{HT^{\frac{3}{2}}}{Zn}. \tag{2}$$

Putting as an average for the photosphere and chromosphere $H = 20$ gauss, $T = 10^4$ degrees, and $Z = 1$, we obtain

$$\omega_e \tau_e = 2.10^{13}/n. \tag{3}$$

Hence when n, the number of positive ions per cm.^{3}, decreases below 10^{14} cm.$^{-3}$, the conductivity becomes anisotropic. In the chromosphere the density is 10^{15} cm.$^{-3}$ at the boundary to the photosphere and 10^{10} cm.$^{-3}$ in the upper chromosphere. Consequently the anisotropy begins in the lower part of the chromosphere. In the corona ($T = 10^6$ degrees) we have

$$\omega_e \tau_e = 2.10^{16}/n. \tag{4}$$

† Owing to the low degree of ionization in the photosphere, n_e in formula 3.23 (66) is so small that in the upper photosphere the second term in the same formula becomes larger than the first, but only a few times.

For $n = 10^8$ we have $\omega_e \tau_e = 2.10^8$, so that the parallel conductivity is more than 10^{16} times the cross-conductivity. It is doubtful whether this very high figure really has any physical significance.

As has been pointed out in § 3.24, an electric current due to a certain electromotive force is not usually determined by the conductivity but by the inductance of the circuit.

5.22. *Solar magnetic fields.* The solar magnetic fields are of fundamental importance to electromagnetic phenomena in the sun. Hale (1908) discovered that Zeeman effect occurs in the Fraunhofer lines from sunspots, from which he concluded the existence of strong magnetic fields. These are of the order 10^2–10^3 gauss; values up to 3,000–4,000 gauss are often observed. All sunspots are associated with magnetic fields.

Later Hale, Seares, van Maanen, and Ellerman (1913) reported the existence of a general magnetic field, also discovered by Zeeman effect measurements. The field was found to be of a dipole character. Spectral lines deriving from different heights in the photosphere seemed to give different values of the field. Hence they concluded that the field decreased very rapidly with the height in the photosphere, although this is incompatible with the dipole character. The strength at the poles varied from 50 gauss in deep photospheric layers to 10 gauss higher up. Such weak fields give a Zeeman effect of the same order as the errors in the measurements, so the values must necessarily be very uncertain.

The investigation has given rise to much discussion. Theories of the rapid decrease of the field in the photosphere have been proposed, but Cowling (1929) has shown that they are not valid. The rapid decrease seems to be very unlikely from a theoretical point of view. The measurements are so uncertain that the decrease with height cannot be considered as certain. Within the errors of measurement the results would be compatible with a dipole field with a polar strength of about 25 gauss. Some authors seem inclined to deny the existence of a field of this order of magnitude. Recently the Zeeman effect measurements have been repeated by Thiessen (1946). Like the earlier authors he reported at first a field with a polar strength of about 50 gauss (his value was 53 ± 12 gauss). Repeated measurements have not confirmed this result, and he now claims that the field, if any, must be below 8 gauss.

Hale and his collaborators also reported that the axis of the magnetic field was inclined at an angle of 6° to the rotational axis. As will be shown in § 5.24, it is from a theoretical point of view very likely that the magnetic and rotational axis coincide. The observational evidence in favour of an inclination is not definite.

The uncertainty of the results of Zeeman effect measurements makes it very important to consider other arguments concerning the general field. More or less direct and definite arguments are supplied by the study of the following phenomena:

1. The progression of the sunspot zone from high latitudes towards the equator is according to the magneto-hydrodynamic theory of sunspots (see § 5.3) compatible with a dipole field with a polar strength of between 10 and 40 gauss.
2. The fact that the low-energy part of cosmic radiation seems to be less intense than was expected has been interpreted as an effect of the solar magnetic field, which outside the earth's orbit is still strong enough to affect cosmic rays (see § 7.22). No definite conclusion about the solar magnetic field has as yet been reached in this way.
3. In the theory of magnetic storms and aurora as developed in Chapter VI, the existence of a general field is essential. The order of magnitude could be the above, but one order of magnitude lower or higher would not be ruled out.
4. Recently Babcock (1947) has reported the existence of very strong general magnetic fields in stars. For 78 Virginis he finds 1,500 and for BD–18°3789 (HD 125248) 5,500 gauss. If stars have such strong magnetic fields, it is quite reasonable that the sun should have a field of about the strength reported by Hale.

Taking all the observational and theoretical evidence together, the most likely properties of the field are the following. The field at the surface and above it is a dipole field with the *polar strength of about* 25 *gauss*, the axis coinciding with the rotational axis. There are no definite arguments for the view that the field decreases more rapidly than a dipole field, nor that it makes an angle with the axis.

The sign of the field is the same as that of the terrestrial field.

We know very little of the magnetic field in the sun's interior. Like all magnetic fields, it must be due to electric currents of some kind. Elsasser (1939) has analysed the different possible causes of such currents and has concluded that the only possibility is that they are produced by thermo-electric or electrochemical forces. He has developed a theory on the thermo-electrical basis. Although the theory is extremely interesting it leads to difficulties which seem to be rather serious.

Cowling (1945*a*) has pointed out that the time constant of the solar field is probably very large (10^{10} years) and that consequently the solar field may be a decaying relic of a primeval state. From this assumption

he is able to derive the shape of the field in the sun's interior (Fig. 5.3). As he points out, his results are valid only if no hydrodynamic motion changes the shape of the field.

Attempts to account for the solar magnetic field by introducing a new law of nature have been discussed in § 1.1.

If the solar field is a dipole field at the surface, it is likely to be approximately a dipole field even to some depth below the surface. It cannot

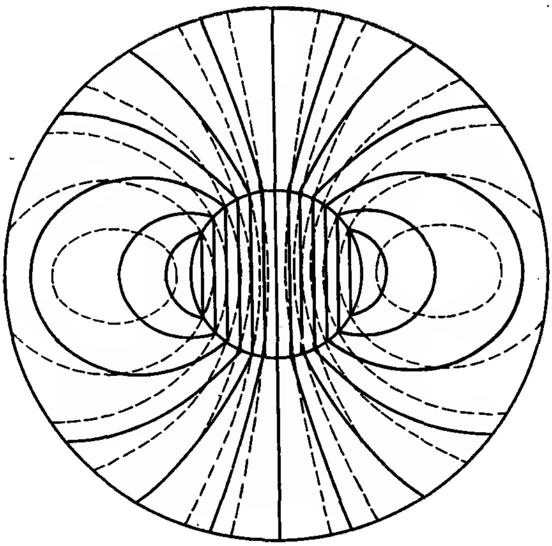

FIG. 5.3. Assumed magnetic field compared with Cowling's decaying field.

have a dipole character up to the centre, because a dipole field becomes infinite at the dipole itself (which is assumed to be situated at the centre). In the central region the field is likely to be more or less homogeneous. Hence the mathematically simplest approximation to the real field would be a homogeneous field from the centre up to a certain central distance R_1 and a dipole field outside R_1. This implies that the current producing the field is uniformly distributed over the thin spherical shell R_1, flowing along its latitude circles, which of course could be only a very rough approximation. If a is the sun's dipole moment, the strength, H_h, of the homogeneous field of this model is given by

$$H_h = 2a/R_1^3.$$

This field is parallel to the axis. Outside the sphere R_1 the dipole field is defined by

$$H_R = \frac{2a}{R^3}\sin\varphi,$$

$$H_\varphi = -\frac{a}{R^3}\cos\varphi,$$

where R is the central distance and φ the latitude (compare § 1.2).

The magneto-hydrodynamic theory of sunspots (§ 5.3) is compatible with the above model, although it would be too much to say that it actually supports it. The value of R_1 can be estimated in three different ways: the progression of the sunspot zone gives $R_1 < 2{\cdot}4.10^{10}$ cm.; from the shape of the sunspot whirls the value $R_1 = 1{\cdot}3\text{–}1{\cdot}6.10^{10}$ cm. is derived; the length of the sunspot cycle gives $R_1 = 1{\cdot}8\text{–}2{\cdot}0.10^{10}$ cm. A survey of the different models is given in Table 5.3 and Fig. 5.3.

TABLE 5.3

Model		*Ratio* $\frac{\textit{central field}}{\textit{polar field}}$	*Remarks*
Homogeneous-field-dipole-field model	$R_1 = 1{\cdot}3\text{–}1{\cdot}6.10^{10}$ cm.	80–150	From shape of sunspot whirls
	$R_1 = 1{\cdot}8\text{–}2{\cdot}0.10^{10}$ cm.	40–60	From length of sunspot cycle
Cowling's decaying field	. . . .	40	—

In the following we adopt the value $H_h = 2{,}000$ gauss, which corresponds to $H_h/H_p = 80$ and $R_1 = 1{\cdot}6.10^{10}$ cm.

5.23. *Magneto-hydrodynamic waves in the sun.* As the sun possesses a general magnetic field and as solar matter has good conductivity, it is obvious that the conditions for magneto-hydrodynamic waves are favourable. The sun is highly anisotropic to the waves. They travel along the magnetic lines of force, every line of force oscillating almost entirely independently of the others (only the continuity of the hydrodynamic motion must be satisfied). The wave velocity

$$V = H/\sqrt{(4\pi\rho)}. \tag{5}$$

The field being approximately homogeneous, H_h, for $R < R_1$ and a dipole field for $R > R_1$ (compare § 5.22), we have for the *homogeneous part* $V = H_h/\sqrt{(4\pi\rho)}$. For the *dipole part* the field at the magnetic axis is given by

$$H_a = 2a/R^3 = H_p(R_0/R)^3. \tag{6}$$

Then we have for the velocity V_a along the axis

$$V_a(R) = a/[R^3\sqrt{(\pi\rho)}], \tag{7}$$

where ρ is a function of R. The values of V_a are given in Table 5.1 and Fig. 5.1. In the major part of the sun's interior the velocity is of the order 1 m./sec. At a point (R, φ), where φ is the latitude, the magnetic field $H = \frac{1}{2}H_a\sqrt{(1+3\sin^2\varphi)}$ and hence the velocity

$$V = \tfrac{1}{2}V_a\sqrt{(1+3\sin^2\varphi)}. \tag{8}$$

The velocity component in the radial direction is, because of equations 1.2 (9) and 1.2 (5), given by

$$V_R = V\cos\alpha = V_a\sin\varphi = V_a(R)\sqrt{(1-R/r_e)}, \tag{9}$$

where $r_e = R/\cos^2\varphi$ is the central distance of the point where the line of force in question intersects the equatorial plane (compare equation 1.2 (6)). The tangential velocity

$$V_\varphi = R\,d\varphi/dt = V\sin\alpha = \tfrac{1}{2}V_a\cos\varphi, \tag{10}$$

which gives
$$\frac{d\varphi}{dt} = \frac{V_a(R)}{2r_e\cos\varphi}. \tag{11}$$

When passing through the sun the waves are damped in three ways: (1) 'Joule damping' due to the finite electrical conductivity, which converts wave energy into Joule heat; (2) viscous damping, due to the viscosity of the medium; and (3) 'gravitational damping', due to the work done in mixing adjacent layers of different entropy.

The Joule damping can be estimated from equation 4.32 (25):

$$z_0 = \frac{\mu^{\frac{3}{2}}}{\sqrt{\pi}}\frac{\sigma}{c^2}\frac{H_0^3}{\rho^{\frac{3}{2}}}\frac{1}{\omega^2}, \tag{12}$$

where z_0 is the distance in which the amplitude decreases to $1/e$, μ $(= 1)$ is the permeability, σ/c^2 the conductivity, H_0 the sun's general magnetic field, ρ the density, and ω the angular frequency of the waves. The values of $\omega^2 z_0$ for the sun's interior are given in Table 5.1 and Fig. 5.1, where Cowling's values of the electrical conductivity have been used (see § 5.21). For the outer layers of the sun we assume according to § 5.1 that these consist of almost pure hydrogen. Hence we put the density

$$\rho = nm_{\mathrm{H}}$$

(n = number of atoms cm.$^{-3}$, m_{H} = mass of hydrogen atom). If the current associated with the wave flows perpendicularly to the magnetic field we should use the cross-conductivity. Introducing 3.23 (66) and putting

$$n_e = \eta n; \qquad Z = 1,$$

we obtain
$$z_0 = H_0^3\,\omega^{-2}(b_1 H_0^2\,\eta^{-2}n^{-\frac{1}{2}}T^{\frac{3}{2}}+b_2\,n^{\frac{1}{2}}T^{-\frac{3}{2}})^{-1}, \tag{13}$$

with $b_1 = 2{\cdot}4 . 10^{-10}$ cm.$^{-3}$ g.$^{\frac{1}{2}}$ sec. grad$^{-\frac{3}{2}}$; $b_2 = 2{\cdot}4 . 10^{-22}$ cm.^{2} sec.$^{-1}$ g.$^{\frac{3}{2}}$ grad$^{\frac{3}{2}}$.

The function $\omega^2 z_0$ is given in Table 5.2 and Fig. 5.2.

The viscous damping has not been calculated but is probably not very important in the sun.

The 'gravitational damping' is according to § 4.8 positive when the waves pass through a stable part of the sun, i.e. where the entropy increases outwards so that work is done by the waves in mixing adjacent layers. If the entropy decreases outwards so that the solar atmosphere is unstable, the damping is negative, resulting in an increase in amplitude. It must be observed that the gravitational damping occurs only when the material velocity (**v**) of the waves has a vertical component. Waves with horizontal **v** are unaffected.

Besides the damping the wave vectors **v** and **h** may change because the solar magnetic field H_0 and the density ρ change. When the relative change of H_0 and ρ over one wave-length is small, the energy of the wave remains unchanged. According to § 4.7, equations (1) and (2), v is proportional to $\rho^{-\frac{1}{4}}$, and h is proportional to $\rho^{+\frac{1}{4}}$. On the other hand, if H_0 or ρ varies rapidly, this is equivalent to a rapid change in refractive index and may cause a partial reflection of the waves.

When a wave travels in a medium with variable refractive index, reflection occurs if the variation is large over the distance of one wave-length. The refractive index is inversely proportional to the wave velocity. Consequently if V varies rapidly, reflection of the magneto-hydrodynamic waves takes place. Rydbeck (1948) has treated the reflection of waves (magneto-hydrodynamic and electromagnetic) which travel in the z-direction in a medium whose refractive index is a function of z. His formulae make it possible to calculate the reflection for cases when the refractive index can be approximated to certain functions. As we are mainly interested in the order of magnitude, it may be sufficient to use his result for a medium where the density falls exponentially in the z-direction with the scale-height ζ_0. In this case no considerable reflection occurs for wave-lengths shorter than

$$\lambda_c = 4\pi\zeta_0. \tag{14}$$

Hence the *critical angular frequency* (lowest frequency transmitted) is

$$\omega_c = \frac{2\pi V}{\lambda_c} = \frac{H_0}{4\sqrt{(\pi\rho)}\zeta_0}. \tag{15}$$

Suppose that in a certain layer of the sun the density ρ decreases as

$$\rho = \rho_0 \exp(-R/\zeta_1) \tag{16}$$

(R = central distance). Then the scale-height is ζ_1. When the magnetic

lines of force are vertical we have $\zeta_0 = \zeta_1$. When they make an angle α with the vertical, the distance which a wave must travel in order to reach a layer where the density has decreased to $1/e$ is given by

$$\zeta_0 = \frac{\zeta_1}{\cos\alpha}. \tag{17}$$

As according to § 1.2, equation (9), we have $\cos\alpha = 2\sin\varphi(1+3\sin^2\varphi)^{-\frac{1}{2}}$, where φ is the latitude, and, as $H = \frac{1}{2}H_a(1+3\sin\varphi)^{\frac{1}{2}}$, where H_a is the field at the axis, we obtain for the critical frequency at latitude φ

$$\omega_c^{(\varphi)} = \omega_c^{(a)}\sin\varphi, \tag{18}$$

with
$$\omega_c^{(a)} = \frac{H_a}{4\sqrt{(\pi\rho)}\zeta_1} = \frac{V_a}{2\zeta_1}. \tag{19}$$

The function $\omega_c^{(a)}$ is given in Tables 5.1 and 5.2 and Figs. 5.1 and 5.2.

Frequencies a few times smaller than ω_c are reflected; frequencies a few times higher than ω_c are transmitted without considerable reflection.

5.24. *The non-uniform rotation.* Observations of the motion of sunspots and other markings of the photosphere indicate that different latitudes rotate with different angular velocities. These results might be due to systematic motion of the markings in relation to the main mass of matter at the latitude in question, but Doppler effect measurements confirm that the angular velocity really varies with the latitude φ. Several formulae have been proposed for the angular velocity Ω. One of the best is probably

$$\Omega = \Omega_a+\Omega'\cos^2\varphi = \Omega_e-\Omega'\sin^2\varphi. \tag{20}$$

Newton (1934) has given values for Ω_e and Ω', derived from sunspot observations for the epoch 1878–1933. As mean values we can put

$$\Omega_e = 14{\cdot}4^\circ/\text{day}; \qquad \Omega_a = 11{\cdot}6^\circ/\text{day}; \qquad \Omega' = 2{\cdot}8^\circ/\text{day}. \tag{21}$$

Measurements of the Doppler effect in the chromosphere indicate that the angular velocity increases with the height over the photosphere.

No observations of the rotation of layers below the photosphere can be made, but a theorem of Ferraro (1937) makes it possible to calculate the average state of rotation, provided that we know the shape of the magnetic field. According to Ferraro every point of a magnetic line of force must rotate with the same average angular velocity, because otherwise very strong electric currents would be induced. The applicability of the theorem is a consequence of the good conductivity of solar matter. It could also be interpreted in the following way. Owing to the good conductivity the magnetic lines of force are 'frozen' into solar matter. If the angular velocity of a certain part of a magnetic line of force differs

from that of other parts, the line becomes twisted. A twisted line of force tries to straighten itself by sending out a magneto-hydrodynamic wave. Although during a short time different parts of a line of force may move with different velocities in connexion with the passage of a magneto-hydrodynamic wave, the average over a sufficiently long time must give the same angular velocity for all parts of the line of force. The time which is necessary to establish the same average angular velocity of all parts of a line of force is given by the time of travel of the magneto-hydrodynamic waves.

As, according to observations, different latitudes of the photosphere rotate with different angular velocities, the same must hold for the interior of the sun. The surface traced out by a magnetic line of force rotating around the solar axis may be called an *isorotational surface*, because all points of this surface must rotate with the same angular velocity.

Assuming a magnetic dipole field, a line of force may be defined by the parameter r_e, given as the central distance of its intersection with the equatorial plane (see § 1.2). The equation of a line of force being

$$R = r_e \cos^2\varphi,$$

we find from (20) that the angular velocity is given by

$$\Omega = \Omega_a + R_0 \Omega'/r_e, \tag{22}$$

R_0 being the solar radius. Only the values for $r_e \geqslant R_0$ are observationally derived. If the field in the centre is homogeneous out to the sphere R_1, where it goes over into the dipole field, a line of force defined by the parameter r_e intersects the sphere R_1 at the point (R_1, φ_1), with

$$\cos^2\varphi_1 = R_1/r_e.$$

Hence in the homogeneous core the distance from the axis of the line of force

$$P = R_1 \cos\varphi_1 = R_1^{\frac{3}{2}} r_e^{-\frac{1}{2}}, \tag{23}$$

from which follows

$$\Omega = \Omega_a + \Omega' R_0 R_1^{-3} P^2. \tag{24}$$

We do not know whether the formula holds for $R_1 > P > R_1^{\frac{3}{2}} R_0^{-\frac{1}{2}}$, because the corresponding lines of force do not intersect the surface.

Based upon these considerations a calculation of the sun's angular momentum has been made by Lundquist (1948).

Suppose that we have a conducting sphere with a magnetic field which is symmetrical around an axis. According to Ferraro's theorem differential motions of the isorotational surfaces defined by this magnetic field are allowed. Hence a differential rotation around the magnetic axis is

possible. If we then rotate the whole body around an arbitrary axis, this motion is of course also allowed from the electrodynamic point of view. Thus we may say that the magnetic axis must coincide with the axis of the differential non-uniform rotation, but not necessarily with the axis of the resultant rotation. Dynamical effects probably make all the three axes coincide. The observational result claimed by Hale and

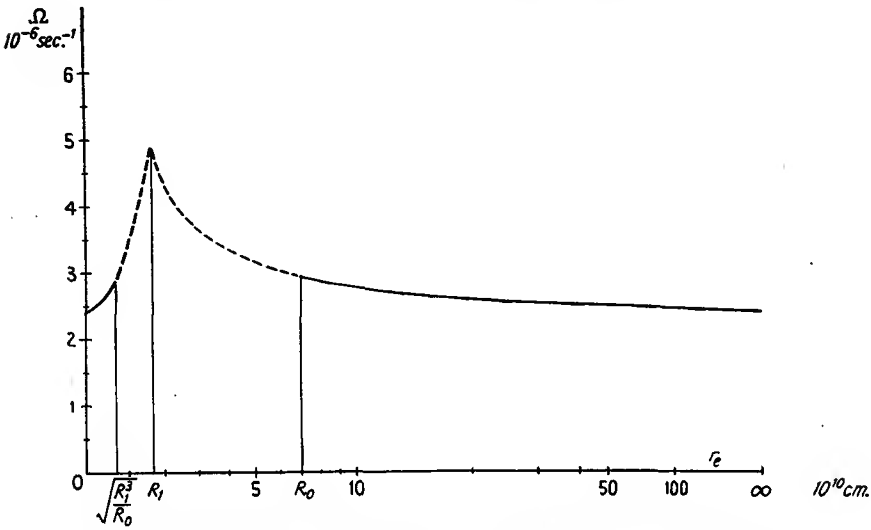

FIG. 5.4. Angular velocity in the sun's equatorial plane, from (22) and (24).

his collaborators (1913) that the magnetic axis is inclined 6° in relation to the rotational axis cannot be considered as definite.

The observed increase in velocity above the photosphere is very difficult to interpret theoretically. Unless the electric conductivity or the vertical component of the magnetic field is *many* powers of 10 less than is reasonable for other reasons, no appreciable increase of the angular velocity with the height seems to be possible as a stationary state. A check of the observations would be desirable. It is of interest to note that, according to Waldmeier (1946), the corona rotates with about the same angular velocity as the photosphere.

5.25. *Survey of electromagnetic phenomena in the sun.* Any turbulence in the sun must produce magneto-hydrodynamic waves. As stated in § 5.1 there are at least two turbulent regions, one in the core and the other in the photosphere.

The turbulence in the core is necessary from a theoretical point of view because the nuclear energy production depends on a much higher power

(16th or 18th) of the temperature than the radiation. Hence the energy production occurs in a very small region near the centre, and the radiation does not suffice to carry away the energy. Convection must take place in order to transmit the energy outwards. According to current stellar models the sun can be divided into two parts: a 'convective core' from the centre out to about 1.10^{10} cm. and a 'non-convective' region outside this limit. In the latter the transfer of solar energy outwards is effected by radiation only and the entropy increases outwards. In the convective core the entropy is slightly decreasing outwards so that the conditions become unstable and convection sets in.

The discovery of the magneto-hydrodynamic waves may change our conception of the conditions in the sun's interior. In the 'convective core' convective motions, as all motions, must give rise to magneto-hydrodynamic waves. In fact, it is necessary to treat the convection in terms of magneto-hydrodynamic whirls increasing their velocity according to the pattern of § 4.8. A large fraction of the energy of convection is converted into magneto-hydrodynamic wave energy.

As Walén (1944) first demonstrated, the damping of the waves due to causes other than gravitation is likely to be negligible in the sun's interior. As is seen from Table 5.1 and Fig. 5.1, $\omega^2 z_0$ has a value of about 300 in the region $3.10^{10} < R < 5{\cdot}5.10^{10}$. In order to pass this region without considerable damping the waves must have $z_0 > 3.10^{10}$ cm. This gives $\omega < 10^{-4}$, corresponding to a period $T = 2\pi\omega^{-1}$ of about one day. Waves with longer periods are not damped very much. Because of the large size of the 'convective core' it is likely that waves with long periods are generated. These travel outwards, producing convection even in the 'non-convective' region. In fact the waves transfer every state of motion from the core to the outer layers. As these are stable (entropy increasing outwards), the waves are subject to 'gravitational damping', and dissipate their energy in mixing adjacent layers.

Consequently it is not appropriate to speak of a 'convective' and a 'non-convective' region, because both are more or less convective. Instead we may distinguish between a 'stable' outer region, where the entropy increases outwards, and an 'unstable' inner region, where the entropy decreases outwards. It is unlikely that these regions coincide with the 'non-convective' and 'convective' regions, because the convection forced upon the stable region by the waves may change the temperature gradient calculated according to current ideas of stellar structure. A revision of the stellar model may be necessary. On the other hand, it is reasonable for the border between the stable and

unstable regions to be situated not very far from the non-convective–convective border of current stellar models. According to Blanch, Lowan, Marshak, and Bethe (1941) this border is situated at $0{\cdot}84 . 10^{10}$ cm. from the solar centre.

The magneto-hydrodynamic waves produced by the convection in the solar core travel outwards along the magnetic lines of force until they reach the photosphere. In the 'magneto-hydrodynamic theory' of sunspots the magnetic fields associated with the waves are identified with the strong magnetic fields always observed in sunspots, the spots themselves being considered as secondary effects of these fields. The theory requires that the waves should have the shape of magneto-hydrodynamic rings, such as treated in § 4.6. When a ring intersects the solar surface, a bipolar sunspot is caused.

The magneto-hydrodynamic theory of sunspots is discussed in the following sections. In § 5.31 the travel of the waves in the sun's general magnetic field is discussed. Some conclusions regarding the field can be drawn from the study of the progression of the sunspot zones from high latitudes towards the equator. In § 5.32 the shape of the whirl rings is constructed. In § 5.33 an outline of a general theory of the generation of the sunspot whirls is given.

The other turbulence region of the sun is situated in the photosphere or not very far below it. The turbulence manifests itself as the *granulation*. When the solar disc is observed with high magnification, it appears as consisting of dark and light grains, 'granulae', which appear, change their shape, and disappear with a period of a few minutes. Their average size is 10^8 cm.

The granulation has been explained theoretically as an effect due to the ionization of hydrogen. As the temperature increases downwards, the hydrogen becomes completely ionized, which means that the average molecular weight of the solar gas decreases. Instability is caused, which produces convection.

It must be observed that the turbulence of the photosphere need not necessarily derive from an instability located in the photosphere itself. The instability may be situated in a deeper layer in which it produces a turbulence, which is transmitted to the photosphere by means of magneto-hydrodynamic waves. But, wherever its place of generation may be, the turbulence observed in the photosphere must be associated with magneto-hydrodynamic waves, which must travel outwards along the magnetic lines of force. When they reach the corona the decrease of the conductivity causes a damping of the waves, so that the wave energy

is transformed to heat. The fact that the corona is at a temperature of about 10^6 degrees, 100 times hotter than the photosphere, is accounted for by the heating due to the damping of the waves (see § 5.4).

Of the waves generated by the granulation it is only the highest frequencies which reach the corona. The lower frequencies are reflected in the upper chromosphere because of the rapid change in refractive index. The 'sunspot waves' from the convection zone in the core contain only low frequencies (because, as we have seen, frequencies $\omega > 10^{-4}$ are damped) and are already reflected considerably below the photosphere (according to Fig. 5.2 below $n = 10^{18}$ particles cm.$^{-3}$). At the reflection a multitude of electromagnetic phenomena occur. As long as a magneto-hydrodynamic ring travels in a homogeneous medium, the disturbance is confined to the ring itself. The centrifugal force and the contraction of the magnetic lines of force balance each other exactly, and the electric polarization is compensated by the induced electric field. When the ring reaches, say, the photosphere, where the density varies rapidly, the compensations are no longer complete. Hence not only in the ring itself but also in its environment electromagnetic disturbances occur. Mighty electromotive forces and strong currents are produced in the solar atmosphere, causing phenomena of the kind of electric 'discharges'. It may be possible that many of the 'solar activity' phenomena can be identified with effects produced more or less directly by the magneto-hydrodynamic waves reaching the surface. In fact there is no doubt that several aspects of the 'solar activity' are of electromagnetic character. In particular, the magnetic field of sunspots is certainly of primary importance to the occurrence of the spots, and the recently discovered phenomenon, called 'solar noise', is of course also of electromagnetic origin. The electromagnetic origin of prominences and solar flares is less obvious. These phenomena are usually attributed to some mechanism working below the visible layers in the sun, but most theories are hardly more than a complicated way of expressing our ignorance of their ultimate cause. In this treatise an attempt will be made to account for most of the solar activity as due to electromagnetic phenomena, more or less directly produced by disturbances originating in the convection zones of the sun. It is clearly understood that this can only be a preliminary attack on the extremely difficult problem.

5.3. Sunspots

Sunspots are regions of the photosphere which have a lower temperature (about 4,600° K.) than normal (5,700° K.). They occur periodically

with a cycle of about eleven years. Every cycle starts with spots at a latitude of about 30° (north and south). Each spot does not change its latitude very much, but the zones where the spots occur are slowly displaced towards the equator. When the zones have reached 10° or 15°, the number of spots attains a maximum. Before a spot cycle fades out near the equator, a new cycle has started at $\pm 30°$.

Every spot is associated with a strong magnetic field (up to 4,000 gauss). The spots usually occur in pairs. The two components of a pair have about the same latitude and always opposite magnetic polarities. During a certain cycle all pairs on the northern hemisphere have the same polarity sequence, e.g. the westward (preceding) spot is a magnetic north pole and the eastward (follower) spot a south pole, and on the southern hemisphere they have at the same time the opposite sequence. For both hemispheres the polarity sequence of a certain cycle is opposite to that of the preceding cycle.

Most theories of sunspots have tried to find a more or less close analogy with the cyclones in the terrestrial atmosphere or in any case treated them as hydrodynamic whirls. The most developed theory is due to Bjerknes (1926), who correlates the sunspots with the non-uniform rotation which, according to him, should produce vortices, which are identified with spots.

The hydrodynamic-whirl theories of sunspots encounter the difficulty that frequently no pronounced vorticity is really observed. A still more serious drawback is that they cannot account for the existence of the spot magnetic fields. The magnetic properties are probably essential, because no sunspot without a strong magnetic field has ever been observed. It seems very difficult to find along these lines a mechanism producing the magnetic field, for, even if a spot is considered as a whirl, it cannot be understood how a whirl could produce a magnetic field of the observed magnitude. (If the matter of a whirl has an electrostatic charge a current is certainly produced, but its magnetic field is *many* orders of magnitude too low.)

The spot magnetic fields stress the difference between sunspots and terrestrial cyclones. In fact dynamically the solar atmosphere must behave altogether differently from the terrestrial atmosphere, because the latter is an electrical insulator whereas the former is a good conductor. A motion in (the lower parts of) the terrestrial atmosphere does not produce any electromagnetic effects, whereas in the solar atmosphere magneto-hydrodynamic waves are produced which in general transmit the state of motion in the direction parallel and antiparallel to the

magnetic field. This should be carefully observed before making analogies between solar and terrestrial phenomena.

An attempt to introduce electromagnetic phenomena in the discussion of sunspots has been made in the *magneto-hydrodynamic theory* of sunspots (Alfvén, 1943*a*, *b*; 1945*a*, *b*; 1948*a*; Walén, 1944), which will be reviewed here. The main postulates of the theory are the following:

The primary photospheric phenomenon of a sunspot is the magnetic field. All other properties, e.g. the decrease in temperature, should be regarded as secondary effects of the magnetic field. According to observations all sunspots are associated with strong magnetic fields. When occasionally a strong magnetic field is observed in a spot-free area, a spot usually occurs there very soon.

The spot magnetic fields are not actually generated in the photosphere but received as magneto-hydrodynamic waves from the sun's interior. The waves start from the solar core and proceed along the magnetic lines of force. All spots of a certain cycle are due to the same set of disturbances in the core, but, as the waves reach high latitudes of the surface earlier than the equator, the spot zone starts at high latitudes and proceeds towards the equator.

The initial cause of sunspots is the convection in the solar core, which after magneto-hydrodynamic transmission to the surface causes spots. The periodic occurrence of sunspots reveals that the convection in the core proceeds discontinuously.

5.31. *Progression of the sunspot zone.* For the development of the theory and the comparison with observations we shall start with a discussion of the transmission of the waves from the solar core to the surface and show how the geometry of the solar magnetic field is connected with the progression of the sunspot zone from high latitudes (30–40°) to the equator.

Let us assume, provisionally, that the solar magnetic field is a dipole field and that a sudden magneto-hydrodynamic pulse occurs near the central dipole. This gives rise to waves which travel along the lines of force, reaching the solar surface after a time $T(\alpha)$, which is a function of the latitude α. This time, according to 5.23 (9) and 5.23 (7), is given by

$$T(\alpha) = \int\limits_0^{R_0} \frac{dR}{V_R} = \frac{\sqrt{\pi}}{a} \int\limits_0^{R_0} \frac{\sqrt{\rho} R^3 \, dR}{\sqrt{(1-R/r_e)}}, \tag{1}$$

where r_e and α are connected [compare § 1.2 equation (6)] by

$$R_0 = r_e \cos^2\alpha,$$

a is the magnetic dipole moment, and ρ is density. At the surface the disturbance gives rise to sunspots. Thus a single pulse would give rise to sunspots at a certain latitude only at that moment when the wave front reached the latitude in question. If we knew the time of the pulse at the centre and observed the time when the spots occurred at a certain latitude, we could find $T(\alpha)$. As there seems to be no possibility of observing the time of a pulse at the centre, we can only

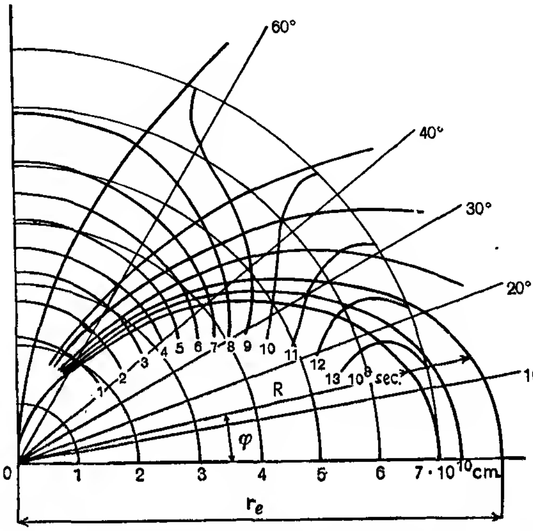

FIG. 5.5. Successive positions of the wave front from a pulse at the centre of a dipole field. (After Walén.)

hope to find the dependence of $T(\alpha)$ upon the latitude, but not its absolute value.

In reality, however, the problem is more difficult, for there is always a series of pulses, constituting a sunspot cycle, which give rise to a lengthy series of spots at each latitude. In order to find the function $T(\alpha)$ we might take a certain characteristic event of a sunspot period (still related to a certain latitude), e.g. the beginning, maximum, or end of the spottedness at the latitude in question. None of these, however, is usually very well defined. We therefore prefer to take the weighted mean $T_1(\alpha)$ of time t,

$$T_1(\alpha) = \int t\,dS \Big/ \int dS \tag{2}$$

(S = spottedness as defined later), of the spots at a certain latitude as a measure of the time when the wave front reaches the latitude.

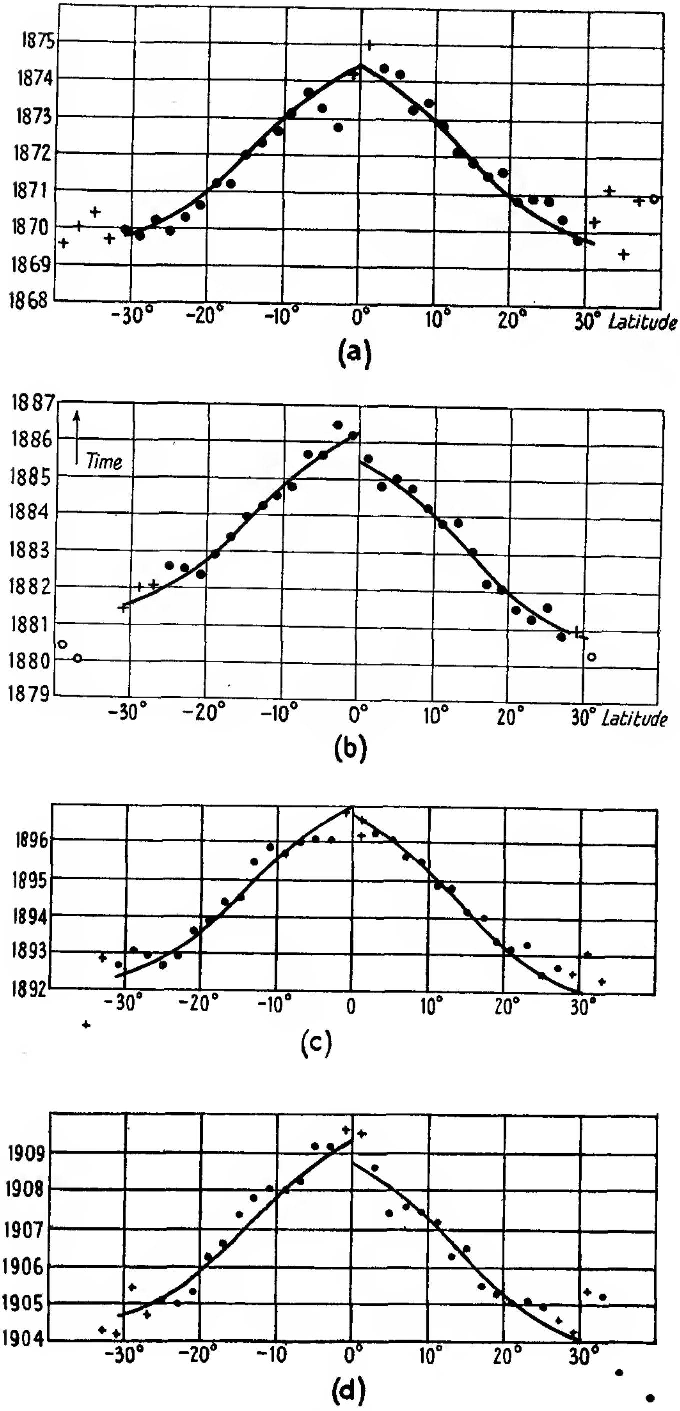

FIG. 5.6 *a-d*. Sunspot progression curves. Cycles 11, 12, 13, 14.

According to our assumptions the time of transit $T(\alpha)$ is given by

$$T(\alpha) = T_1(\alpha) - T_0, \tag{3}$$

where T_0 is the time of the disturbance at the centre.

The function $T_1(\alpha)$ can be derived from sunspot observations. We can

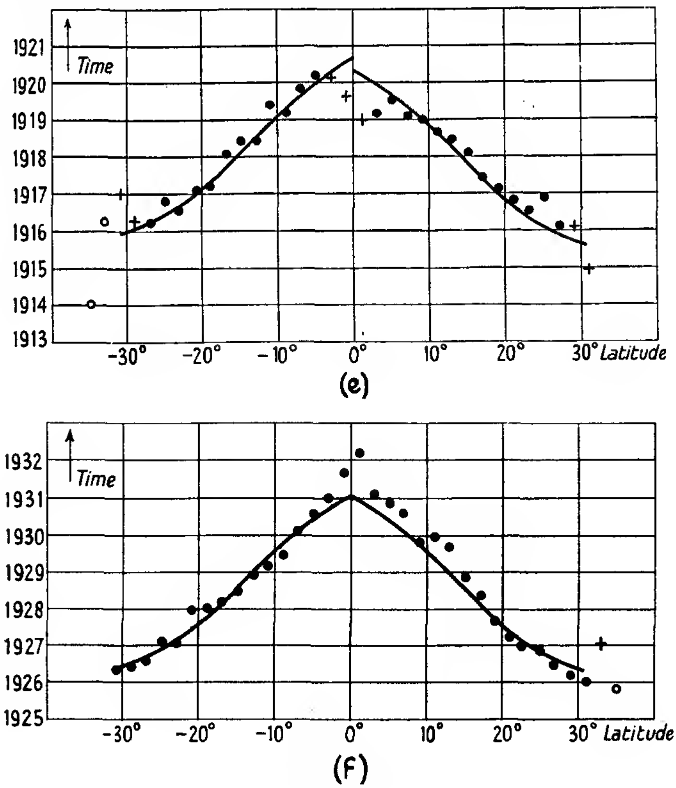

FIG. 5.6 *e, f*. Sunspot progression curves. Cycles 15, 16.

let S represent either the number of spots or the spotted area at the latitude α. The first alternative gives much more reproducible results (see Alfvén, 1945*a*) so we choose that. The function $T_1(\alpha)$ has been calculated for each hemisphere separately for the sunspot cycles 11–16, comprising the time 1868–1932. The curves are shown in Fig. 5.6, in which the full lines represent the average curve for both hemispheres and all periods. This curve is shown separately in Fig. 5.7. It is evident that the function is rather well reproduced during all the cycles. In some cases the curve for the northern hemisphere seems to be displaced in relation to that of the southern hemisphere, so that one gets the

impression that the two hemispheres are to some extent independent of each other. In the mean values the difference between the hemispheres is small. The northern curve may be a little steeper than the southern one, but the difference is only a few per cent.

It is of interest to note that the slope of the sunspot progression curve shows a velocity of 5–10° per year, whereas the displacement of the mean latitude of spots varies between 4° and 1° per year (see Waldmeier, 1941, p. 127).

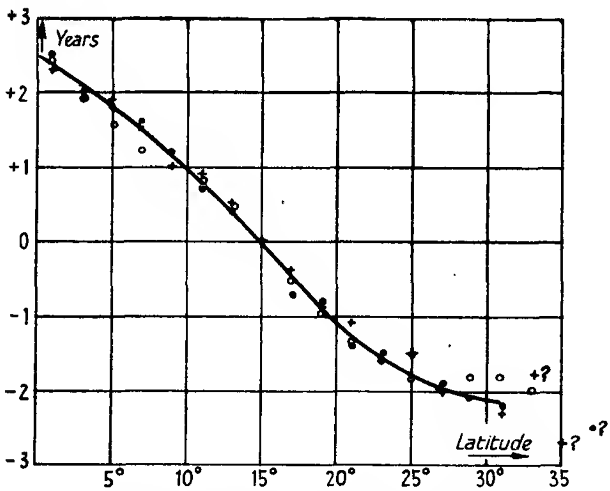

FIG. 5.7. Sunspot progression curves. Means. ● cycle 11-12; ○ cycle 13-14; + cycle 15-16.

The time of transit $T(\alpha)$ can be calculated from the density and the magnetic field in the sun's interior. According to (3), $T(\alpha)$ will have the same shape as $T_1(\alpha)$. The calculation of $T(\alpha)$ can be made by numerical integration. The result shows that the shape of the curve becomes about the same whether we choose the density according to the polytrope model or the Blanch–Lowan–Marshak–Bethe model, only the absolute value is about 1·4 times larger in the latter case.

The agreement between the observational curve and $T(\alpha)$ as calculated from (1) is not good. This could hardly be expected because the solar magnetic field could not possibly be a dipole field up to the centre. As pointed out in § 5.22, the magnetic field is likely to be fairly homogeneous near the centre. If so, the lines of force cutting different parts of the solar surface do not converge to a single point but are separated even in the core. Thus the start of the waves becomes

arbitrary. The simplest assumption is that all the waves start at the same time from the equatorial plane.

Further it must be observed that a sunspot is not a purely photospheric phenomenon; it has certainly its 'roots' far below the surface. Hence the appearance of a sunspot is not correlated with the time when the wave reaches exactly the photosphere but rather a some-

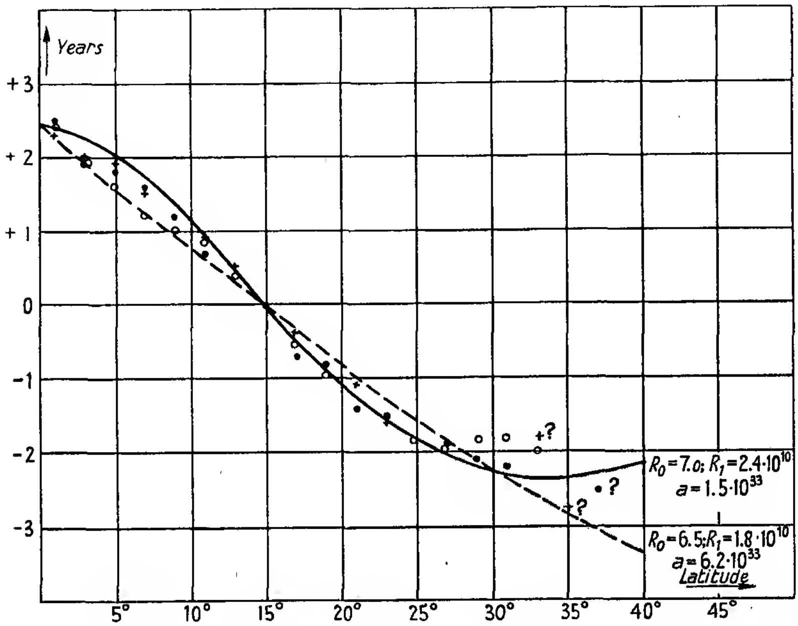

Fig. 5.8. Comparison between theoretical curves and observation.

what deeper layer. In the paper cited calculations have been made of the time a wave needs to travel out from the equatorial plane of the homogeneous core to the photosphere ($R_0 = 7{\cdot}0\,.\,10^{10}$ cm.) and to a depth of $0{\cdot}5\,.\,10^{10}$ cm. ($R_0 = 6{\cdot}5\,.\,10^{10}$ cm.) below the photosphere. In Fig. 5.8 the curves corresponding to $R_0 = 7{\cdot}0\,.\,10^{10}$ and $R_0 = 6{\cdot}5\,.\,10^{10}$ cm. are plotted. In both cases the values of a and R_1 (boundary between dipole field and homogeneous core) are chosen so as to make the absolute value and the average slope of the curves agree with the observational values. If we confine ourselves to stating that the observational curve is intermediate between the two theoretical curves, we may conclude that

$$1{\cdot}5\,.\,10^{33} < a < 6{\cdot}2\,.\,10^{33} \text{ gauss cm.}^3,$$

corresponding to a polar strength

$$9 < H_p < 37 \text{ gauss},$$

and further that $$R_1 < 2{\cdot}4 . 10^{10} \text{ cm}. \tag{4}$$

Because of the number of assumptions we have made, the results are of course not very definite. Anyhow they give some hint of the possible conditions in the sun's interior.

5.32. *Shape of magneto-hydrodynamic whirl rings.* The fact that sunspots usually occur in pairs with opposite magnetic polarities indicates that they are produced by magneto-hydrodynamic whirl rings. Hence Walén (1944) has especially studied the case of a magneto-hydrodynamic wave having the form of a ring. Some of his results have already been quoted in § 4.6.

In order to derive the actual shape of a ring from sunspot observations we shall here discuss its behaviour when it reaches the solar surface (see Alfvén, 1945 *b*). The discussion of the details of the photospheric phenomena will be reserved for § 5.37.

We expect a bipolar sunspot to be due to the arrival at the surface of a magneto-hydrodynamic ring. Each element of the ring is transmitted with the wave-velocity V, $V = H(4\pi\rho)^{-\frac{1}{2}}$. When the ring is cut by the solar surface one spot occurs at each of the intersections. If no secondary phenomena existed, we should expect that when the ring at first reached the solar surface two sunspots of different polarities would be created very close together. When the ring continued to move outwards, the two spots representing the intersections between the ring and the solar surface would increase their mutual distance until a maximum equal to the diameter of the ring was reached. Thereafter the distance would diminish, and theoretically we should expect the spots to move very close together again and disappear when the last part of the ring reached the solar surface. In reality the last phase is seldom observed, which may be considered as an effect of secondary phenomena, but at least in some large bipolar spots the major part of the above process can be discerned. This makes it possible to *construct the shape of the whirl ring* from the observations of the behaviour of a bipolar sunspot.

During the motion from the core outwards the whirl rings are deformed because of the changes in the magnetic field and the velocity. It is of special interest to study the shape of a whirl ring when it is situated in the central region, where it is supposed to be created. We assume as earlier that the sun's magnetic field can be regarded as approximating to a dipole field outside a certain central distance R_1 and to a homogeneous

field inside R_1 (Fig. 5.9). As the diameters of the rings, as judged from the distance between the two spots of a bipolar group, are usually less than one-fifth of the solar radius when the rings are close to the surface, it is easily seen that even when they are moving in the homogeneous core the dimensions of the rings, or at least of their projection upon the equatorial plane, are small in comparison to R_1 (one-fifth or less).

Suppose that a magneto-hydrodynamic ring gives rise to a sunspot, which at the time t is situated at latitude α and longitude λ, and at another (earlier) time t' at α' and λ'. Suppose further that the time necessary for

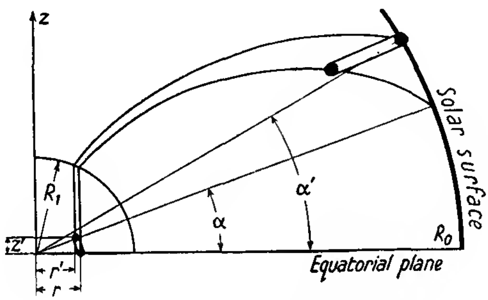

FIG. 5.9. Motion of a whirl ring from the core to the surface.

a wave to travel from the equatorial plane to the latitude α (or α') equals T (or T'). At the instant $t-T$ that part of the ring which causes the spot at (t, α, λ) was situated in the equatorial plane (in the homogeneous core). Employing cylindrical coordinates (r, λ, z) and observing that a magnetic line of force in the dipole field is given by $R^3r^{-2} =$ const. (where $R^2 = r^2+z^2$), we find that its situation is defined as follows:

$$r = r_0 \cos\alpha \quad \text{with} \quad r_0 = (R_1^3/R_0)^{\frac{1}{2}}, \tag{5}$$

$$\left.\begin{aligned} \lambda &= \lambda, \\ z &= 0. \end{aligned}\right\} \tag{6}$$

At the same instant the ring element (t', α', λ') has the coordinates

$$\left.\begin{aligned} r' &= r_0 \cos\alpha', \\ \lambda' &= \lambda', \\ z' &= \int_0^\tau V\, d\tau, \end{aligned}\right\} \tag{7}$$

with
$$\tau = t-T-(t'-T'). \tag{8}$$

V represents the wave velocity along the line of force which passes (parallel to the solar axis) through the point $(r', \lambda', 0)$. In the cases of

interest to us, z' is small, so that the mass density ρ is almost constant. Thus we can write approximately

$$z' = V(t-t'+T'-T). \tag{9}$$

According to § 5.31 the difference $T'-T$ equals the difference in latitude divided by the velocity of the sunspot zone progression. From Fig. 5.7 it is obvious that this drift U is roughly constant, having the value $U = 2{\cdot}0.10^{-7}$ heliographic degrees/sec. Thus, we have

$$T'-T = (\alpha-\alpha')/U, \tag{10}$$

and consequently $$z' = V[t-t'+(\alpha-\alpha')/U]. \tag{11}$$

If we place a Cartesian coordinate system (x', y', z') with the origin at $(r, \lambda, 0)$, the x-axis pointing outwards in the direction of the solar radius, we have

$$\left.\begin{aligned} x' &= r'\cos(\lambda'-\lambda)-r = r_0\{\cos\alpha'\cos(\lambda'-\lambda)-\cos\alpha\} \\ y' &= r'\sin(\lambda'-\lambda) = r_0\cos\alpha'\sin(\lambda'-\lambda) \\ z' &= V\{t-t'+(\alpha-\alpha')/U\} \end{aligned}\right\}, \tag{12}$$

or approximately (because $\lambda'-\lambda$ and $\alpha'-\alpha$ are small)

$$\left.\begin{aligned} x'/r_0 &= (\alpha-\alpha')\sin\alpha \\ y'/r_0 &= (\lambda'-\lambda)\cos\alpha \\ z'/V &= t-t'+k(\alpha-\alpha') \end{aligned}\right\}, \tag{13}$$

where $k = 1/U = 5.10^6$ sec./heliographic degree $= 58$ days/degree.

In order to construct the original whirl ring according to the above equations, we must know the whole life of a bipolar sunspot. As examples, unusually great spots, observed by Waldmeier, and statistical material, resulting from Greenwich observations, have been treated. The result is seen in Fig. 5.10.

In the diagrams the scale of the z-axis in relation to the x- and y-axes is arbitrary since we do not know the values of r_0 and V.

The shape of a hydrodynamic whirl is usually almost circular. It is reasonable to assume that the original whirls starting the magneto-hydrodynamic waves in the solar core have this shape. If this is the case it is obvious that the plane of a whirl in the core is almost parallel to the solar axis. This follows immediately for those spots which have no latitude displacement, because for them x is always zero, but even for a spot with such large latitude displacement as Waldmeier's spot 'a', the whirl plane in the core is almost perpendicular to the equatorial plane. In fact, in the xy-plane the ratio between the axes of the curve, approximated to an ellipse, is 1:8. So that, if this ellipse is the projection

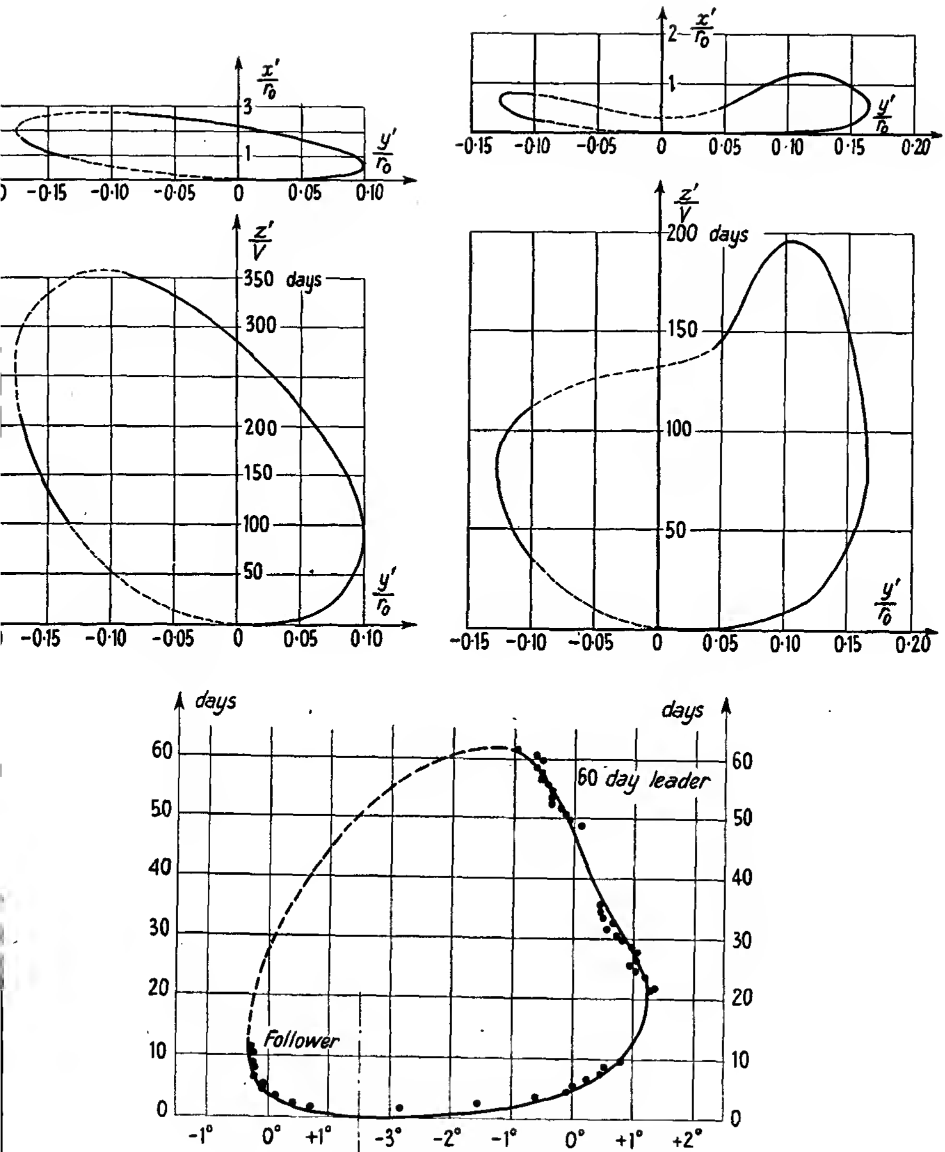

FIG. 5.10. Construction of whirls from sunspot data.

Above: Waldmeier's spots '*a*' and '*b*'.

Below: Statistical material from Greenwich observations.

of a circular whirl, the plane of this whirl is inclined $\cos^{-1}\frac{1}{8}$ ($= 83°$) to the equatorial plane. Thus it seems legitimate to conclude that *the planes of the whirl rings are usually almost perpendicular to the equatorial plane* (and perpendicular to the meridian plane) when they are close to the equatorial plane in the central parts of the sun.

The behaviour of a whirl intersecting the solar surface and creating a sunspot there is illustrated in Fig. 5.11.

Introducing the condition that the whirl is almost circular when near

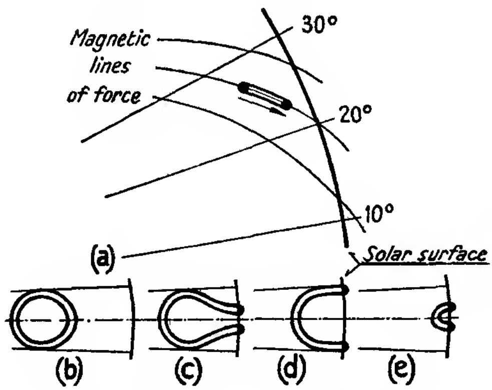

FIG. 5.11. Whirl ring approaching the surface.

the equatorial plane, we obtain a relation between r_0 and V. Denoting the greatest difference in the y' and z' values by Δy and Δz, and putting

$$\Delta y/r_0 = A; \qquad \Delta z/V = B, \tag{14}$$

we can determine A and B from the diagrams. If the assumption that the whirl is approximately circular is expressed by the simple condition $\Delta y = \Delta z$, we obtain

$$Ar_0 = \Delta y = \Delta z = BV, \tag{15}$$

or, in consequence of (5),

$$R_1^{\frac{3}{2}} R_0^{-\frac{1}{2}} V^{-1} = BA^{-1}. \tag{16}$$

If the sun's dipole moment is a, we have in the homogeneous central field $H = 2aR_1^{-3}$ and

$$V = \frac{a}{\sqrt{(\pi\rho)}R_1^3}, \tag{17}$$

$$R_1^{\frac{9}{2}} = \frac{a\sqrt{R_0}}{\sqrt{(\pi\rho)}}\frac{B}{A}. \tag{18}$$

Putting $a = 4{\cdot}2.10^{33}$ gauss cm.^{3} and $\rho = 80$ gm. cm.$^{-3}$, we obtain ($R_0 = 7.10^{10}$ cm.)

$$R_1^{\frac{9}{2}} = 7.10^{37}B/A. \tag{19}$$

From Fig. 5.10 we obtain:

TABLE 5.4

		Waldmeier		
		'a'	'b'	*Greenwich*
A	degrees	15·5°	17°	7°
	radians	0·27	0·30	0·12
B	days	360	190	65
	sec.	31.10^6	16.10^6	$5{\cdot}6.10^6$
B/A	sec.	$1{\cdot}2.10^8$	$0{\cdot}55.10^8$	$0{\cdot}47.10^6$
R_1	cm.	$1{\cdot}6.10^{10}$	$1{\cdot}35.10^{10}$	$1{\cdot}30.10^{10}$
H	gauss	2100	3400	3800
V	cm./sec.	65	110	120

Consequently, using the assumptions that the original whirl is almost circular and that the solar magnetic field has a homogeneous core out to R_1, we find that R_1 is about $1{\cdot}4.10^{10}$ cm. This is compatible with the result obtained in § 5.31, viz. $R_1 < 2{\cdot}4.10^{10}$ cm.

According to the diagrams the deviations from the circular form are considerable, although not very large. They may be due to secondary effects, in part at the solar surface but perhaps also in part at the birthplace of the whirls.

It is reasonable to attach a certain importance to the place where the rings, deforming themselves during the transmission, are circular. The assumption that the rings are created circular in the core has given the above values of the field in the core. It may be of interest to observe that as, according to Fig. 5.1, V_a is almost constant in the interior of the sun out to $R = 5.10^{10}$ cm., so the rings are circular during most of their passage through the sun.

5.33. *Generation mechanism and the sunspot cycle.* In the preceding paragraph we have found the shape and orientation of the whirl rings which produce sunspots when reaching the photosphere. The great problem now arises where and how they are produced. Combined with this is the question why the production of whirl rings exhibits the eleven-year periodicity.

The mechanism proposed here is rather speculative, and the only reason why it is given is that it leads to a conclusion (concerning an expected correlation of sunspot distributions) which seems to be confirmed observationally. The initial cause of sunspots is supposed to be convection in *two* activity regions in the solar core, the co-operation between them resulting in a sort of relaxation oscillation. A disturbance from one of the regions travels to the other, where it initiates another disturbance, which is sent back to the first region, where it triggers a new

disturbance. The sunspot period is given by the time of travel of magneto-hydrodynamic waves between the two regions. Too much importance should not of course be given to the details of this tentative model.

A whirl producing spots at E (see Fig. 5.12) must have been generated somewhere on the magnetic line of force $EFGHJKJ'G'F'E'$. This line passes stable as well as unstable parts of the sun (compare § 5.25). In the photosphere, eventually extending to some depth below it, there is an unstable region where convection may set in, thus producing magneto-hydrodynamic waves. The large size of the sunspot rings makes it unlikely that they are produced in this rather thin unstable region. Moreover, the periodicity, the latitude dependence, and the progression of the spot zones would probably be difficult to explain.

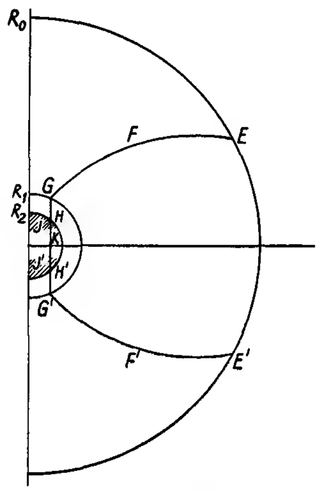

Fig. 5.12. Line of force in sun. Shaded areas represent the supposed 'activity regions'.

Below the unstable surface region there is a vast stable region extending to the unstable core. The border between the unstable core and the stable region outside it is probably a sphere which has a radius (R_2) of about 10^{10} cm. (compare § 5.25). It is not necessary that R_2 should equal R_1 (limit between homogeneous magnetic field and dipole field). There is some indication that R_1 should be somewhat larger than R_2 (see §§ 5.22 and 5.32), but this is not essential for our discussion.

In a stable region a whirl could only be damped. Hence the only place where the whirl rings could be generated is in the unstable core. Observing that (as found in § 5.32) in the core the plane of the whirls is perpendicular to the equatorial as well as to the meridian planes, we can restrict the place of generation still farther. When a whirl with this orientation is located at K near the equatorial plane the gravitation acts perpendicularly to the plane of the whirl. Under this condition the quantity Θ_z in 4.8 (23) is zero and the whirl cannot increase. It is evident that convection cannot produce a whirl in a plane perpendicular to the force of gravitation. Hence the neighbourhood of the equatorial plane is excluded as the place of generation of the actually observed spot rings.

The only possibility which is left is that the whirls are generated in two regions J and J', one on each side of the equatorial plane, inside the

unstable core. In these regions the gravitation has a component in the direction of the magnetic field, so that a whirl could increase by means of the mechanism treated in § 4.8. The quantity Θ_z, which is zero in the equatorial plane and at the border to the stable region, exhibits one maximum (near J and J') in each hemisphere.

Our result that the whirls are not generated near the equatorial plane but in two regions on both sides of it is very important just because it introduces the concept of *two* regions of generation ('activity regions'). This gives a new line of approach to the understanding of the sunspot period. This length of time may be connected in some way with the time of travel of magneto-hydrodynamic waves from one activity region to the other.

According to Walén (1944) an initial hydrodynamic whirl produces two magneto-hydrodynamic waves, one moving in the direction $+\mathbf{H}$ and the other in the direction $-\mathbf{H}$. If a whirl, moving as a wave, for example, in the positive H-direction, increases in velocity by the amount Δv during the small time Δt, this is equivalent to adding to the initial wave a new whirl. This new whirl will go out not only in the positive H-direction (as the increase of the initial whirl) but also in the negative H-direction. Thus every increase of a wave gives rise to a 'recoil' wave transmitted in the opposite direction. It is easily seen that the recoil whirl due to the increase of a certain whirl has the *same direction of rotation* as the initial whirl and the opposite sense of magnetic field.

In the same way a decrease in the whirl velocity is equivalent to adding a whirl of the opposite sense of rotation. Consequently if a whirl is damped a recoil whirl is produced, but with the *opposite direction of rotation* and the same sense of magnetic field.

There is a fundamental analogy between a magneto-hydrodynamic wave travelling along magnetic lines of force and an ordinary electromagnetic wave along a transmission line. If at a certain point a damping resistance is parallel-connected to the transmission line, a partial reflection of the waves is caused. In the same way the damping of magneto-hydrodynamic waves causes a partial reflection. The increase of a magneto-hydrodynamic wave in an unstable region corresponds to the parallel-connexion of a negative resistance to the transmission line.

If the increase in velocity is continuous, as found in § 4.8 under idealized conditions, a long continuous recoil whirl is sent out, the velocity of which is proportional to the rate of increase dv/dt of the initial whirl. In reality the increase is probably more discontinuous, because the density (or entropy) is irregular in the unstable core, which

means that a series of discrete whirls are sent out. The average strength of these whirls is proportional to the average of dv/dt.

Suppose that a ring passes, for example, the northern activity region J, producing recoil whirls there, and later reaches the northern spot zone E. The recoil whirls ultimately reach the southern spot zone E'. It is easily found that they produce bipolar spots with the *same* polarity sequence as the original whirl produces in the northern spot zone.

Waves produced in, for example, the northern activity region must of

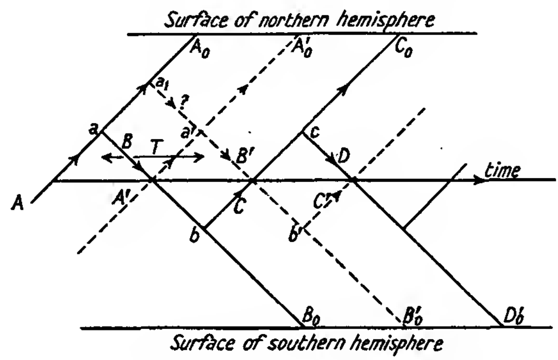

FIG. 5.13. Generation of sunspot waves.

course reach the northern spot zone earlier than the southern spot zone. As the spots produced by a certain original whirl have the same polarities on both hemispheres (according to statements above), they must belong to consecutive sunspot cycles. This means that we must expect the waves generated at the northern maximum of Θ_z to reach the southern spot zone eleven years after they have reached the northern one. In the same way, whirls generated at the southern maximum of Θ_z near J' reach the southern spot zone eleven years earlier than the northern. Consequently *the time necessary for a wave to travel between the northern and the southern activity region must be eleven years.*

A possible mechanism for the generation of the waves would be the following. Suppose that a weak magneto-hydrodynamic whirl (with its plane parallel to the magnetic field) starts at A somewhat to the south of the equatorial plane in the sun's unstable core and proceeds in the northward direction (Fig. 5.13). As long as the whirl velocity $v \ll V$ (wave velocity) the increase in velocity is rather small (see § 4.8). Consequently the 'recoil whirls' are faint. The whirl increases however, and when at a certain time t it has reached a velocity v equal to V, the rate of increase becomes rapid, which means that considerable recoil whirls

are sent out. In the neighbourhood of the northern maximum of θ_z the rate of increase is especially rapid, and consequently the recoil whirls especially strong. The wave proceeds farther into the stable region where it becomes somewhat damped, and finally it reaches the solar surface in the northern spot zone, where it produces sunspots at A_0.

The recoil whirls proceed southward and increase exponentially until their velocity v has reached the value V, which occurs at a time which we call $t+T$. Because of the exponential increase, T depends upon the original strength of the recoil whirl. Weak whirls will leave the unstable region before having reached the velocity V, producing only weak secondary recoil whirls. Because of this only the strongest of the recoil whirls are of interest. Restricting our discussion to the strongest recoil whirls, the generation of sunspot waves may go on according to Fig. 5.13. An original whirl A reaches the critical velocity, $v = V$, at a. A recoil whirl B (finally producing spots at B_0) is sent out. When B reaches the critical value at b, a new recoil whirl (CcC_0) is produced. The time T from a to b depends upon the original strength of the recoil whirl produced at a and upon the value of Θ_z.

If the original whirl rotates in the positive direction, all the whirls considered do so, and all spots produced by them have the same polarity sequence. Therefore only every odd sunspot cycle at one hemisphere and every even cycle at the other hemisphere could be accounted for. The other cycles must be due to an analogous series of whirls ($A'B'C'$...) with the opposite sense of rotation and intermediate between those of the first group. As the direction of rotation does not affect the growth of the whirls, both types of whirls are equally possible.

The mechanism outlined may explain some essential facts, but it is not altogether satisfying. Certainly, as both of the two series of waves develop under the same conditions, it is reasonable that they should become symmetrical (having the same form and amplitude). There remains to be explained, however, why all the whirls of one series have the same sense of rotation. Probably we must assume that there is some sort of coupling between all the whirls of a certain cycle. This coupling may be effected hydrodynamically by some general flow within the solar core, but this is hardly more than a guess.

We must also look for an explanation why there are two series of waves with different senses of rotation. This might simply be the result of some tendency for symmetry, because if there were one series of waves only, the conditions at a certain time would be very asymmetrical. But it may also be explained by the following magneto-hydrodynamic effect

by which whirls of one sense produce whirls of the opposite sense. When a whirl passes the stable region, especially that part (probably rather near to the border of the unstable region) where Θ_z is a minimum (e.g. at a_1 in Fig. 5.13), recoil whirls of the opposite sense are sent out (as shown above). Thus there is a coupling, although perhaps weak, between the two sequences of whirls, and, if initially there is only one type of whirl, the other type is produced automatically.

If the whirls of the second series are exactly intermediate between the whirls of the first series the difference in time between the waves sent out is T (defined above). Thus the sunspot cycle (eleven years) should be equal to T, the time which a whirl needs to increase from a, where it is produced as a recoil whirl, to b, where it is strong enough to produce new recoil whirls itself. As the rate of increase, and hence the tendency for production of recoil whirls, has maxima in the two activity regions, T equals approximately the time of transit between these two regions.

In § 5.32 it was shown, assuming the shape of the whirls in the solar core to be circular, that it was possible to calculate the velocity of the waves in the core. The values obtained (Table 5.4) varied between $V = 65$ and 120 cm. sec.$^{-1}$

It is possible to calculate the velocity independently, using the condition that the time needed for a wave to go from one region of generation to the other should equal eleven years. Both the regions of generation must be situated inside the unstable core, which, according to current solar models (which may need revision!), has a diameter of about 2.10^{10} cm. The distance between the two regions can consequently be estimated to be about $1{\cdot}0$–$1{\cdot}5.10^{10}$ cm. The velocity V needed to cover this distance in eleven years is 29–44 cm./sec. This is somewhat less than the above values but of the same order of magnitude. As the first values depend upon somewhat arbitrary assumptions as to the shape of the magnetic field and the whirls, the agreement can be considered as encouraging. Consequently our result that the 11-year period is determined by the distance between the two active regions is quantitatively consistent.

In § 5.31 the progression of the spot zones was found by calculating the time of travel from the equatorial plane in the core to the surface. This is not incompatible with the concept of the two activity regions, because in the activity regions the pulses are merely amplified. It is quite reasonable that at the moment when the pulses pass the equatorial plane the waves should be symmetrical with respect to this plane, and this is in fact the essence of the assumption in § 5.31.

5.34. *Correlation between intensity and length of the sunspot cycles.* The fact that the intensity of the spot cycles varies indicates that the state in the solar core, where the whirls are generated, is subject to variations. It seems probable that these consist in changes of the instability factor Θ_z. The larger the value of Θ_z the stronger are the waves produced, and the higher the spottedness of the cycle in question.

According to § 5.33, however, a variation of Θ_z means also a variation

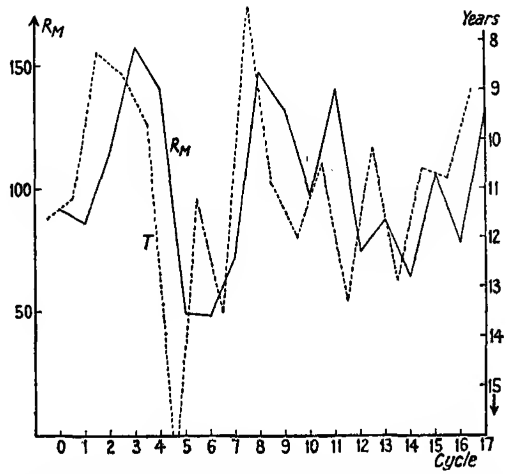

FIG. 5.14. Variation of the maximum relative number R_M (full curve) and the length of the solar cycle (dotted curve).

of the sunspot period T. In fact, theoretically T should be the time which a recoil whirl needs to grow strong enough to produce a new (secondary) recoil whirl itself. This time depends upon the original strength of the primary recoil whirl, which is proportional to the rate of increase dv/dt of the original whirl. As according to equations 4.8 (30) and 4.8 (34) dv/dt is proportional to Θ_z, an increase of Θ_z means an increase of the strength of the recoil whirl and consequently a decrease of T. This time depends also directly upon Θ_z, because a large Θ_z means a rapid increase, so that the critical value ($v \sim V$) is reached more quickly.

Consequently, if Θ_z increases, the sunspot cycle will exhibit a higher intensity and at the same time the length of the sunspot period will be shorter. Hence *theoretically we must expect a negative correlation between the intensity* (expressed, for example, in maximum relative numbers of the cycle) *and the length of the sunspot period.*

In Fig. 5.14 the maximum relative numbers of the sunspot cycles are

represented in the same diagram as the length of the sunspot periods.† It is evident that on the whole an increase of the maximum is accompanied by a decrease of the period. The correlation coefficient has been calculated for the maximum relative number R_n of a cycle and the length of some preceding and following periods (for notation see Fig. 5.15). The result is given in Table 5.5.

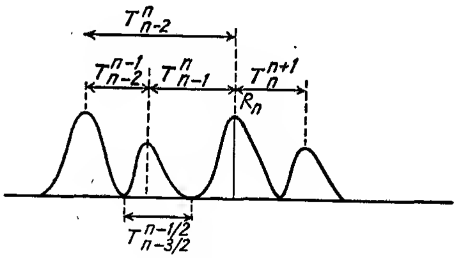

FIG. 5.15.

TABLE 5.5

Correlation Coefficient r between the Maximum Relative Number R_n and the Difference in Time between Maxima (or Minima)

	T^{n-2}_{n-3}	T^{n-1}_{n-2}	T^n_{n-1}	T^{n+1}_n	T^n_{n-2}	$T^{n-\frac{1}{2}}_{n-\frac{3}{2}}$
r	$-0{\cdot}11\pm0{\cdot}23$	$-0{\cdot}35\pm0{\cdot}19$	$-0{\cdot}81\pm0{\cdot}11$	$+0{\cdot}27\pm0{\cdot}23$	$-0{\cdot}85\pm0{\cdot}12$	$-0{\cdot}71\pm0{\cdot}12$

The table shows that whereas there seems to be no significant correlation between R_n and T^{n-2}_{n-3} or T^{n+1}_n, a strong negative correlation exists between R_n and T^n_{n-1}. Even with T^{n-1}_{n-2} there is a negative correlation. Especially strong is the correlation with the preceding double-period T^n_{n-2} and there is also a rather strong correlation with the preceding difference in time $T^{n-\frac{1}{2}}_{n-\frac{3}{2}}$ between consecutive minima.

The correlation between R_n and T supports the general lines of the theory, but it must be admitted that it is not clear why R_n is correlated just with T^{n-1}_{n-2} and T^n_{n-1} but not with T^{n+1}_n.

5.35. *Correlation between spots of different hemispheres and consecutive cycles.* We have found that the production of sunspot waves can be explained as a result of the co-operation of one northern and one southern activity region. The time of travel between these two regions equals approximately the sunspot period (eleven years). The spots occurring, for example, in the northern hemisphere at a certain cycle n have originated in the southern region of generation as 'recoil whirls', produced

† Data taken from Waldmeier, *Ergebnisse d. Sonnenforschung.*

by whirls travelling southwards and later occurring as southern hemisphere spots belonging to the cycle $n-1$. Thus if our theory is correct *it should be possible to find some connexion between the spots of a certain hemisphere and cycle and the spots of the opposite hemisphere and the preceding cycle.*

When a whirl passes one of the activity regions it will in general give rise to a number of recoil whirls. These will increase, travel through the core to the other activity region where they increase still more, and at last, having passed the stable region, where they become more or less damped, they reach the solar surface. Their chance of reaching the surface with energy enough to become visible as sunspots depends, in part at least, upon their initial intensity (intensity of the recoil whirl). Of course the local conditions they have met during their long journey also affect their strength. Anyhow we could expect a statistical correlation between the number of sunspots and the intensity of the recoil whirl production of the primary whirl. The generation of recoil whirls, however, is a result of the increase of the whirl. The more the primary whirl increases in strength, the more recoil whirls it produces.

Consequently, if we observe a very big sunspot in, for example, the northern hemisphere, we should expect that it has increased in strength very much when passing through the northern region of generation. This means that it must have created numerous recoil whirls, which are likely to be visible in the southern hemisphere one cycle later. If the magnetic field is symmetric with respect to the rotational axis, *a big spot at a certain latitude should be associated with an increased activity at the same latitude of the other hemisphere during the following cycle.*

A statistical test of this theoretical result has been made. For reasons developed in detail in the original paper (Alfvén, 1948 *a*) the first six half-years of each of the sunspot cycles 12–17 were chosen. The spotted area of each hemisphere and cycle was normalized and plotted as a function of the latitude. The difference Δ between the functions for the northern and the southern hemisphere was calculated. The functions are plotted in Fig. 5.16, every even cycle with reversed sign. The correlation coefficients between consecutive functions are given in Table 5.6.

The table and the figure indicate that there is a negative correlation between consecutive cycles. A very careful statistical analysis by Galvenius and Wold (1948) demonstrates that the probability that this correlation should be accidental is only 1:200. Thus the correlation can be considered as statistically significant. A special analysis has shown that the correlation cannot be due to a coupling between the spottedness

of a certain hemisphere and cycle and the spottedness of the *same* hemisphere and the following cycle. The final result is that there really *exists a correlation in latitude distribution between an arbitrary cycle of one hemisphere and the following cycle of the opposite hemisphere*, as predicted

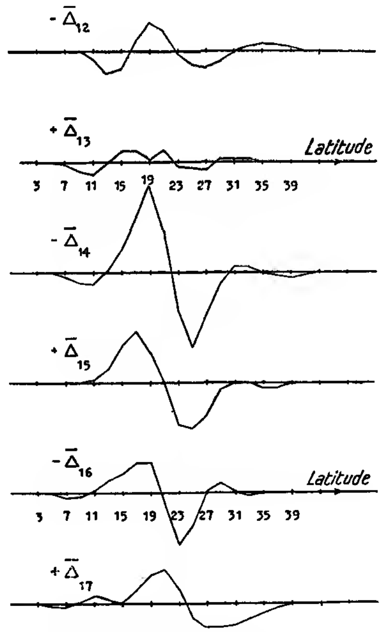

FIG. 5.16. Difference between spotted area on the northern and the southern hemisphere.

TABLE 5.6

Cycle	*Time*	*Correlation between consecutive* $\bar{\Delta}$
12	1879·0–1882·0	
		−0·33
13	1888·5–1892·0	
		−0·68
14	1900·0–1903·5	
		−0·85
15	1912·5–1915·5	
		−0·86
16	1923·5–1926·5	
		−0·12
17	1934·0–1937·0	

by theory. This result is important because it shows that even if many details of the theory may be wrong the concept of the two activity regions in the solar core must be right. It is very difficult to see how the observed correlation could be understood except on this basis.

5.36. *Latitude dependence of spottedness.* If the spottedness, integrated over a whole cycle, is plotted as a function of latitude the diagram of

Fig. 5.17 is obtained, where the number of spots for both hemispheres is combined. The curve rises steeply from near zero at the equator to a maximum between 10° and 15°, from which the spottedness falls again for higher latitudes. The decrease in spottedness on the low latitude side of the maximum can be explained as a simple geometrical effect. Let s be the number of whirls (or the energy) transmitted through

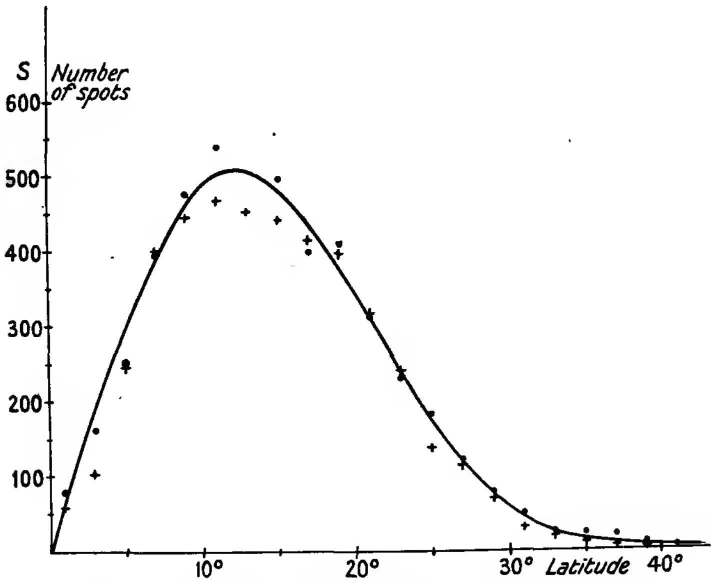

FIG. 5.17. Number of spots as a function of the latitude.
• Cycles 11 and 12. + Cycles 15 and 16.

a unit surface placed in the homogeneous core at right angles to the magnetic field. Introducing the coordinates of § 5.32, we find that the transmission through a differential surface element is $sr\,d\lambda dr$. Projecting upon the solar surface along the magnetic lines of force, and denoting by $S\,d\alpha$ the total transmission received by a ring element comprised between the latitudes α and $\alpha+d\alpha$, we find

$$S = S_0 \sin 2\alpha,$$

with $S_0 = \pi r_0^2 s$. We identify S with the number of spots.

Fig. 5.17 shows S as a function of the latitude α according to observations. In Fig. 5.18 S_0 is plotted as a function of r $(= r_0 \cos\alpha)$. The fact that the latter curve has no maximum shows that the decrease in spottedness in the equatorial region is simply due to the geometry of the dipole field.

The decrease in spottedness as we go to higher latitudes must be explained in some other way. Either the transmission from the

homogeneous core is lower for the lines of force going to high latitudes, or the damping is greater for these lines. (Two early attempts to explain the latitude dependence (Alfvén, 1945*a*, p. 13; and Walén, 1944, p. 45), are obviously unsatisfactory.)

In this connexion an objection by Cowling (1946*b*) against the magneto-hydrodynamic theory of sunspots is of interest. He has pointed out that rings transmitted from the core outwards would never reach the

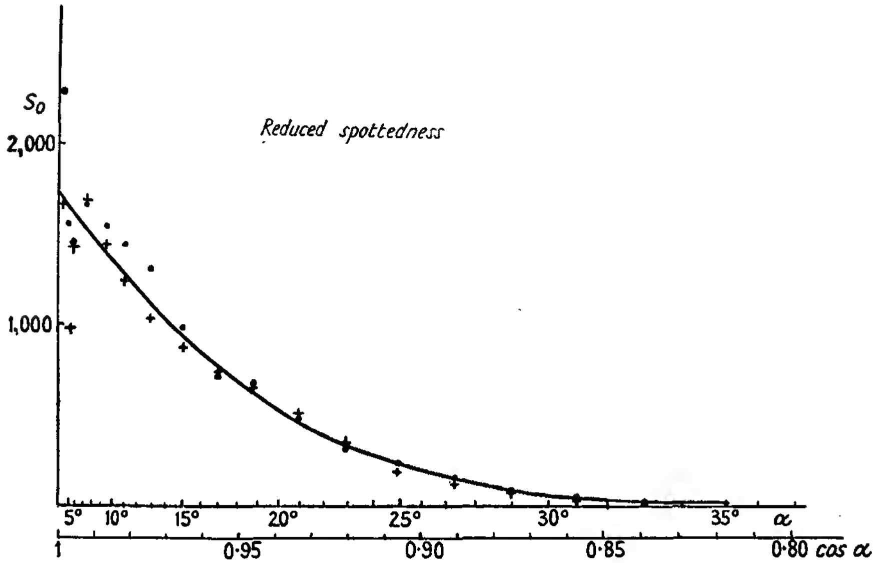

FIG. 5.18. Spot activity S^0 corrected for geometry of magnetic field.

surface because the gravitational damping in the stable layers would rapidly consume all their energy. This objection seems quite reasonable, but, on the other hand, the observational correlation between the two hemispheres, as found in § 5.35, gives strong indication that the waves really do come up from the solar core. Hence it seems necessary to assume that the whirls in some way succeed in avoiding the gravitational damping, for example, by the following mechanism.

Suppose that the angle β between the lines of force and the horizontal plane is rather small. When a whirl climbs along the line of force and reaches a stable layer, it brings heavy matter from a low layer to a higher layer and conversely light matter down to a lower layer. In general, the upper half of the ring will have higher density than the surrounding matter so that it has a tendency to sink down, whereas the lower half of the ring will tend to float up. As the magnetic lines of force are not rigid

the field will be deformed and this deformation will travel upwards as a wave together with the ring (Fig. 5.19). In this way it will be possible for the ring to climb, rotating the whole time in an almost horizontal plane so that the gravitational damping is not very large. The deformation wave is associated with a (slow) vertical motion, which may be reflected or heavily damped near the solar surface, so that it need not affect the surface phenomena very much.

For $\beta = \frac{1}{2}\pi$, the mechanism, sketched above, would not be possible. Instead the whirls would be damped very quickly for the reasons given by Cowling. If β increases from small values the damping increases

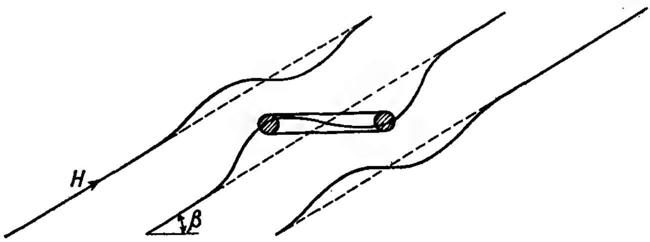

FIG. 5.19. Climbing magneto-hydrodynamic ring.

rapidly. Hence waves may climb from the solar core to the low latitudes of the solar surface along lines of force, which upon the whole make a rather small angle with the horizontal, whereas the lines of force which run to high latitudes make so large an angle that the waves are damped out. This would be a reasonable explanation of the latitude effect. Whether it is the right explanation it is impossible to say at present. It has been given here more to show how important various kinds of secondary effects may be.

(In the unstable core the same effect would have the reverse sense, so that the plane of the whirls would tend to be vertical, thus increasing the acceleration.)

5.37. *Photospheric effects of the whirl rings.* In this section we shall discuss what happens when a magneto-hydrodynamic whirl ring reaches the photosphere. The phenomena are no doubt very complicated. They have been treated by Walén for the general case when the ring makes an arbitrary angle with the magnetic field. As has been shown in § 5.32, the planes of the rings are usually almost parallel to the magnetic field. Hence we shall confine ourselves to that case and start to discuss it under simplified conditions.

Suppose that an incompressible perfectly conducting fluid is limited by a horizontal non-conducting wall and subject to a homogeneous

vertical magnetic field. Suppose further that a circular whirl ring, its plane being vertical, approaches the wall. The problem can be treated (approximately) by the image principle: We assume that the fluid is infinite and that a similar ring with the same sense of h but the opposite sense of v and V is approaching the wall from the other side. When the rings meet and penetrate each other the velocity v through the wall cancels because of the symmetry, but the magnetic fields h amplify each other. The phenomena below the wall do not change very much if the fluid above the wall is taken away. The treatment is approximate because we neglect the non-linear effects which occur when the rings penetrate each other (these vanish only if $h/H \ll 1$, but for

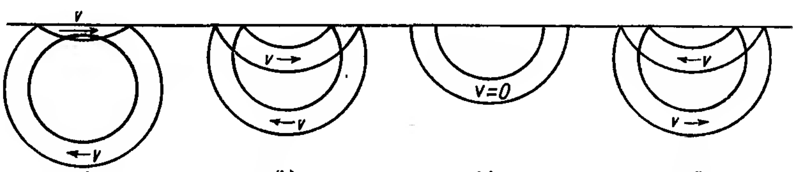

FIG. 5.20. Magneto-hydrodynamic whirl reflected at a surface.

sunspots we have $h/H \approx 100$) and also because the magnetic field above the wall influences to some extent the conditions below the wall. Especially the energy necessary to establish the field above the wall is not accounted for.

In the present state of the theory the phenomena close to the wall during the reflection may be described in the following way.

Just when the ring touches the wall (Fig. 5.20 *a*) a strong horizontal magnetic field ($= h$) appears between the two branches of the ring. At the same time there is a strong hydrodynamic horizontal flow ($= v$) between the branches.

Somewhat later (Fig. 5.20 *b*) two distinct poles have developed. The hydrodynamic flow at the surface is confined to within the surfaces of the poles and decreases rapidly. The magnetic field at the poles becomes more and more vertical. For an infinitely thin ring the vertical component varies as a sine curve. The maximum ($= 2h$) is reached when half of the ring has reached the wall (Fig. 5.20 *c*). At that moment the flow is zero.

When the last half of the ring is reflected the process is repeated in the inverse order, the flow goes up to v and the magnetic field becomes horizontal and equal to h again, after which the surface phenomena end.

It is of interest to discuss the electric field associated with a whirl

ring (compare § 4.6). Consider a straight fixed line perpendicular to the plane of the whirl and situated in such a way that the axis of the ring at a certain moment shall coincide with it. When the torus intersects the line an e.m.f. $(1/c)\int hV\,ds$ is induced because every line element ds of the line is cutting magnetic lines of force at the rate $hV\,ds$. At the same time a polarization e.m.f. equal to vH/c is caused by the hydrodynamic flow. As $v/V = h/H$ and the directions are antiparallel, the resultant e.m.f. is zero. The electromotive forces, which in this ideal case exactly compensate each other, are quite enormous for sunspots. The flux of an ordinary spot, say field = 2,000 gauss, surface = 3.10^{18} cm.^{2}, is 6.10^{21} gauss cm.^{2} Suppose that the torus passes the line considered above in 10^5 sec. Then the e.m.f. equals

$$6.10^{21}/3.10^{10}.10^5 = 2.10^6 \text{ e.s.u.} = 600 \text{ megavolts.}$$

This is the voltage which would be induced in a *fixed* conducting wire placed, for example, immediately above the wall in such a way that it encircles one of the poles in Fig. 5.20 *b*, *c*, or *d*. Below the wall no resultant e.m.f. is caused as long as the flow v gives rise to the compensating polarization.

Starting from the ideal case just discussed we shall now try to approach the real case. As a first step we incline the magnetic field in relation to the vertical. This makes no great change: only the shape of the cross-section of the ring branch is altered by the reflection. If initially circular it becomes elliptic after reflection.

The problem becomes more difficult if we change the incompressible fluid limited by a wall into an atmosphere with continuously decreasing density. Then we have no distinct reflection. The main reflection takes place when the condition $\lambda_c \approx 4\pi\zeta_0$ is satisfied [see 5.23 (14)]. As we have seen in § 5.25, only frequencies below $\omega = 10^{-4}$ cm. pass the region of maximum damping near $R = 4.10^{10}$ cm. Fig. 5.2 shows that such low frequencies are reflected below $n = 10^{18}$ particles cm.$^{-3}$, i.e. below the photosphere. In order to survey the reflection phenomena we ought to treat the case of a whirl in an inhomogeneous medium, or at least the special case of a ring transmitted through a wall separating two liquids of different densities. Unfortunately even this simple case meets mathematical difficulties. Certainly it is easy to treat the problem as long as the magneto-hydrodynamic vectors v and h are parallel to the surface: this case can be reduced to the formulae of ordinary electromagnetic waves. As soon as the vectors have components in the direction of H, however, considerable difficulties arise. In the above discussion of

the reflection of the waves these difficulties appear when accounting for the energy necessary to produce the field above the wall.

5.38. *Pressure effects.* In the photosphere the pressure difference Δp associated with the magnetic field of the magneto-hydrodynamic waves becomes of importance:

$$\Delta p = \frac{1}{8\pi} H^2. \tag{20}$$

For $H = 3{,}000$ gauss the difference in pressure becomes $0{\cdot}36.10^6$ dyn./cm.^{2} = 0·36 atmospheres. This pressure is quite negligible in the interior of the sun, but in the photosphere it is of the same order as the hydrostatic pressure. In fact, if in the photosphere the density $n = 10^{17}$ particles/cm.^{3} and the temperature $T = 6{,}000°$, the gas pressure

$$p = nkT = 10^{17}.1{\cdot}37.10^{-16}.6.10^3$$
$$\approx 10^5 \text{ dyn. cm.}^{-2}$$

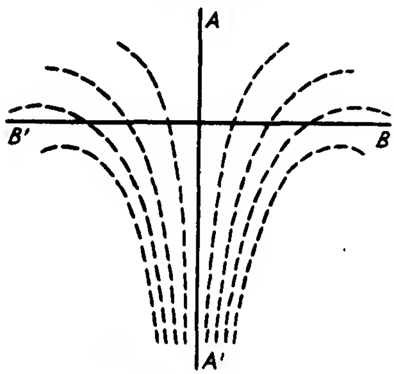

FIG. 5.21. Magnetic field of sunspot.

The magnetic field of a sunspot is likely to be as shown in Fig. 5.21, where BB' is the solar surface and AA' the axis of the sunspot. The magnetic field tends to expand, so that a magnetic pressure Δp according to (20) is directed outwards. This pressure can also be interpreted as due to the force exerted by the magnetic field upon the currents which produce the field. As in a certain layer the sum of the gas pressure and magnetostatic pressure must be constant, the gas pressure in the field region must be reduced. The result is a cooling, but it is still doubtful whether the mechanism is capable of producing the observed amount of cooling.

Cowling (1946) has analysed the conditions critically. He finds that a continuous expansion is necessary in order to keep the spot cool in spite of the heating of radiation. The layer where the coolness of the spot is supposed to start (the 'base' of the spot) must according to his results be situated about 10^9 cm. below the solar surface. Although the diameter of an ordinary spot is of the order of 10^{10} cm., Cowling means that this depth is incompatible with the sharp edge of a sunspot, and he considers this objection to be fatal to any existing theory. Cowling also thinks that the observed magnetic fields are a few times too small to produce the required cooling. Although probably our present knowledge

of the state of the solar atmosphere is not so definite that a discrepancy of less than one order of magnitude must be considered significant, it certainly is very important to look for other processes by which a magnetic field may produce a cooling.

The decrease in pressure due to the magnetic field has also another consequence. Because the lines of force make a large angle with the vertical, the top of the whirl ring 'floats up' because of the reduced pressure in it. When doing so it reaches a region where the solar rotational velocity is smaller. Hence the ring becomes deformed as shown qualitatively in Fig. 5.22. The resulting attenuation of the westward branch

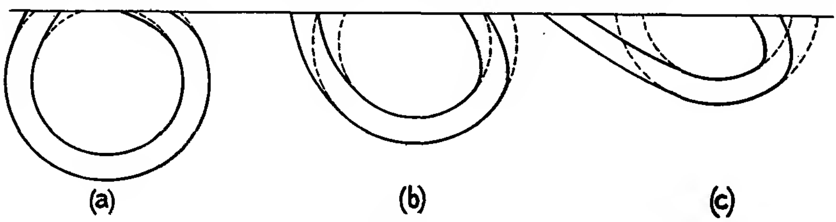

FIG. 5.22. Deformation of whirl ring due to non-uniform rotation.

makes the conditions for the 'follower' less favourable than for the 'leader'. This effect may explain why the leader is usually better developed than the follower.

5.4. The granulation

As pointed out in § 5.25, there are at least two unstable regions in the sun: one near the centre and one in the photosphere. In both the convection produces magneto-hydrodynamic waves. Owing to the large size of the central unstable region, the whirls produced there are very large, and, as we have seen above, these whirls may be the cause of sunspots. The photospheric unstable region is confined to a rather thin layer. Hence the turbulence becomes a small-scale phenomenon, the granulation.

5.41. *Generation of granulation waves.* As any motion of a conducting liquid in the presence of a magnetic field gives rise to magneto-hydrodynamic waves, we must expect the generation of such waves in connexion with the granulation. Their wave velocity

$$V = H_0(4\pi\rho)^{-\frac{1}{2}}. \tag{1}$$

In the photosphere the density varies between 10^{-7} and 10^{-8} g. cm.$^{-3}$ The value of the magnetic field adopted in § 5.22 is 25 gauss at the pole

and 12·5 gauss at the equator. Hence the value of V varies in the photosphere between the limits 7.10^4 and $1{\cdot}1.10^4$ cm. sec.$^{-1}$ As an average we could use

$$V = 3.10^4 \text{ cm. sec.}^{-1} \tag{2}$$

The amplitude of the waves can be estimated in different ways: from the velocities associated with the granulation, and from the difference in brilliance between the granulae. According to some observations the granulae are displaced with a velocity of about 3 km. sec.$^{-1}$ It is possible that this is a real velocity, but maybe the motion is only apparent. If real it corresponds to an average kinetic energy of

$$\tfrac{1}{2}\rho v^2 = \tfrac{1}{2}(3.10^{-8})(3.10^5)^2 = 1{\cdot}4.10^3 \text{ erg cm.}^{-3}$$

The granulation is observed as a difference in brilliance of different small parts of the photosphere. The maximum light difference Δ is about 15 per cent., corresponding to a temperature difference of $\frac{1}{4}\Delta$. If this is caused by adiabatic compression of an ideal gas the change in pressure must be $\frac{5}{2}\times\frac{1}{4}\Delta = \frac{5}{8}\Delta = 9$ per cent. As the pressure in the photosphere is 10^4 to 10^5 dynes cm.$^{-2}$, the pressure fluctuation corresponds to a difference in potential energy of 10^3 to 10^4 erg cm.$^{-3}$ We should expect the turbulence changes in the potential energy to be of the same order of magnitude as those in the kinetic energy. This indicates that the velocity of about 3 km. sec.$^{-1}$, which as we have seen corresponds to about 10^3 erg cm.$^{-3}$, may be real.

In a magneto-hydrodynamic wave the magnetic energy equals the kinetic energy and also the pressure difference Δp associated with the wave:

$$\frac{\mu}{8\pi} h^2 = \tfrac{1}{2}\rho v^2 = \Delta p. \tag{3}$$

Here h means the magnetic field of the wave, which is superimposed upon the general field H_0; and v means the material velocity in the wave. Because the granulation reveals the existence of fluctuations in the kinetic and potential energy of the order of 10^3 to 10^4 erg cm.$^{-3}$, we must expect a varying magnetic field between $h = \sqrt{(8\pi.10^3)} = 160$ and $h = \sqrt{(8\pi.10^4)} = 500$ gauss.

Because of the granulation an irregularly varying magnetic field with an amplitude of a few hundred gauss must, in the photosphere, be superimposed on the general solar magnetic field.

Such fields would give a Zeeman effect broadening of spectral lines of about 0·01 A.U. It would probably be rather difficult though not impossible to find this effect by observation.

The frequency of the granulation waves should be of an order

corresponding to the average life of the granulae, which is a few minutes. This gives $\omega = 0{\cdot}03$ sec.$^{-1}$ Probably the frequencies cover a wide range, as will be discussed below.

5.42. *Transmission of granulation waves.* The magneto-hydrodynamic waves connected with the granulation must be transmitted along the magnetic lines of force. Owing to the finite conductivity they are subject to 'Joule damping' (see §§ 4.32 and 5.23) which converts the wave energy into Joule heat. Waves with material velocities with a vertical component are also subject to gravitational damping (see §§ 4.8 and 5.23), so vertical motions should be more rapidly damped than horizontal motions. In a discussion of the order of magnitude of effects associated with the transmission of the waves we may confine ourselves to the horizontal components. If the turbulent motions are distributed at random the horizontal components should contain two-thirds of the energy.

The transmitted waves are reflected on account of the change in refractive index and damped because of the finite conductivity (see § 5.23). As low frequencies are more easily reflected and high frequencies more damped, these two effects define the limits of the frequency range. In the photosphere the scale height, ζ_1, is about 10^7 cm. and

$$V_a \approx 3.10^4 \text{ cm.},$$

so according to 5.23 (19) the lowest frequency which is not reflected is of the order given below:

$$\omega_c^{(a)} \approx 10^{-3} \text{ sec.}^{-1} \tag{4}$$

On the other hand, we obtain from 5.23 (12) with $\sigma/c^2 = 10^{-9}$, $H_0 = 20$, $\rho = 3.10^{-8}$,

$$z_0 = \frac{10^6}{\omega^2}. \tag{5}$$

In order to avoid damping in the 10^7 cm. thick photosphere we must have $z_0 \gg 10^7$, which gives

$$\omega \ll 0{\cdot}3 \text{ sec.}^{-1} \tag{6}$$

Consequently the waves which could be transmitted through the photosphere must have frequencies in the range

$$10^{-3} < \omega < 0{\cdot}3 \text{ sec.}^{-1} \tag{7}$$

The corresponding period $T = 2\pi\omega^{-1}$.

The average life of the granulae is some minutes, corresponding to a value in the middle of this frequency range.

Going downwards from the photosphere the transmitted frequency

band is displaced towards lower frequencies. At a depth of $0{\cdot}5 . 10^{10}$ cm. ($R = 6{\cdot}5 . 10^{10}$ cm.) we have $\omega^2 z_0 = 10^4$, so that in order to make $z_0 > 10^{10}$ cm. we must have $\omega < 10^{-3}$. Hence all the frequencies transmitted through the photosphere are already damped at this depth.

5.43. *Granulation waves in chromosphere and corona.* We cannot observe the results of the damping of the downward waves. The waves transmitted upwards are of more interest because they are likely to affect observable conditions in the chromosphere and corona. Even if we go upwards the upper limit to the transmitted frequency band decreases when we reach the upper chromosphere and the corona. This means that the waves transmitted from the photosphere are absorbed. The change in the refractive index means that the material velocity and the induced magnetic field of the waves change. As shown in § 4.7, v varies as $\rho^{-\frac{1}{4}}$ and h as $\rho^{\frac{1}{4}}$ (whereas the wave velocity varies as $\rho^{-\frac{1}{2}}$ and, moreover, is proportional to H_0). When the waves move upwards into the chromosphere and the corona, h decreases, and as it is already unobservable in the photosphere, it is likely to be of little importance higher up. The (material) velocity v, on the contrary, increases when ρ decreases.

In the chromosphere, where the density is about 10^{-4} of that of the photosphere, the turbulent velocity should be 10 times as large as in the photosphere or about 30 km. sec.$^{-1}$ In the inner corona, where the density has decreased by 10^{-8}, it should be 100 times as large or 300 km. sec.$^{-1}$ Both values are obtained without taking account of the damping of the waves.

In reality a turbulence velocity of about 20 km. is really observed in the chromosphere. The turbulence in the corona is certainly much less than 300 km. sec.$^{-1}$, in fact only about 20 km. sec.$^{-1}$ Consequently a damping must take place, and this is also what could be expected theoretically, as will be shown below.

A magneto-hydrodynamic wave is always associated with an electric current. A sine wave in the z-direction [see equations 4.31 (11) and 4.31 (13)],

$$h = A \sin \omega\left(t - \frac{z}{V}\right), \tag{8}$$

contains a current with density i,

$$i = A \frac{c\omega}{H_0} \sqrt{\left(\frac{\rho}{4\pi}\right)} \cos \omega\left(t - \frac{z}{V}\right). \tag{9}$$

The current flows in the x-direction when the induced magnetic field h goes in the y-direction and the initial magnetic field H_0 is parallel to the z-axis.

When the conductivity σ is finite, the current produces a Joule heating w_j of the liquid. Its mean for one whole period is given by

$$w_j = \frac{\overline{i^2}}{\sigma} = \tfrac{1}{2}A^2 \frac{c^2}{\sigma} \frac{\omega^2}{H_0^2} \frac{\rho}{4\pi}. \tag{10}$$

As energy is dissipated in this way the waves become damped with the damping exponent given by 4.32 (24). The distance z_0 in which the amplitude of a progressive wave decreases to $1/e$ is given by 4.32 (25) or 5.23 (12). As long as the conductivity is isotropic it is easy to find z_0, but in the chromosphere and corona, where the conductivity is anisotropic, the problem becomes more complicated. For a sine wave [as defined by (8) and (9)], with the current flowing perpendicular to the magnetic field, we should use the cross-conductivity. This leads to 5.23 (13), which function is plotted in Fig. 5.2. It must be observed that the formula is correct only if $h \ll H_0$. For $h \sim H_0$ or $h > H_0$ it may be approximately valid. Further, as in waves started by turbulence the velocities v make arbitrary angles with H_0, the damping is less than in the case of v perpendicular to H_0. Hence equation 5.23 (13) gives a lower limit to z_0, but probably z_0 has the order of magnitude given by the formula.

An accurate calculation of what happens to the granulation waves emitted upwards is at present impossible, because we know too little of the structure of the chromosphere. The corona is at a temperature of about 10^6 degrees (this is discussed in the next section), and it seems likely that it is, at least on an average and approximately, in hydrostatic equilibrium. The conditions of the chromosphere are more difficult to understand. The most puzzling problem of the chromosphere is the low density gradient. It has been generally assumed that the temperature of the chromosphere must be lower than that of the photosphere, but the density gradient corresponds to an atmosphere in hydrostatic equilibrium at about 35,000° [see especially Wildt, (1947)]. Milne (1924) has attempted to explain this by assuming that the chromosphere is supported by radiation pressure, but according to McCrea (1929) this is not possible. Instead McCrea supposes that the intense turbulence which is observed in the chromosphere 'supports' it. The fact that the corona has a very high temperature speaks in favour of a high temperature even in the chromosphere, but there are several arguments against this assumption (see Wildt, 1947).

From an electromagnetic point of view either of the two latter interpretations would be acceptable. The turbulence of the granulation is

transmitted by magneto-hydrodynamic waves to the chromosphere and because of the decrease in pressure the amplitude of the turbulence is amplified (as pointed out above). If McCrea's idea of turbulence support is acceptable from other points of view, it would fit into this picture. On the other hand, if there were no objection to the assumption of a high temperature, this assumption would be simpler. In the absence of a generally accepted theory of the chromosphere, we assume provisionally that the temperature of the chromosphere is 35,000°. The figures of Table 5.2 and Fig. 5.2 are based on this.

As the scale height ζ_s is a measure of the thickness of a layer, we could find the highest frequency ω which penetrates the layer without considerable damping by putting $z_0 = \zeta_s$ in 5.23 (12) or 5.23 (13). The function ω is given in Fig. 5.2. It goes down to the range of frequencies given by (7) in the region of the inner corona ($n = 10^8$–10^7). Hence the frequency band transmitted from the photosphere is converted into heat in this region. As we shall see in the next section, this may be the explanation of the high temperature of the corona.

Against this it may be objected that in the upper chromosphere the low-frequency limit ω_c (given by the reflection of the waves) is so high that rather little of the wave energy really reaches the corona. First it should be observed, however, that the calculation of ω_c is based on simplified assumptions about the chromosphere (constant density gradient, which at the border to the corona suddenly changes to a much lower value). If a smoother curve for the gradient near the limit between the chromosphere and corona is used, the high values of ω_c for $n \simeq 10^9$ disappear. Further, the reflection is calculated on the assumption that the amplitude of the waves is small. In view of the fact that the chromosphere is observed to be violently turbulent, it is doubtful whether reflection really takes place in the predicted way.

As in a wave the kinetic energy w_k equals the magnetic energy w_H, the energy transmitted upwards from the photosphere U is given by

$$U = 2w_H V. \tag{11}$$

As we have found, V is of the order of 5.10^4 cm. sec.$^{-1}$ and $w_H = 10^3$ to 10^4 erg cm.$^{-3}$ Thus we have

$$U = 10^8\text{–}10^9 \text{ erg cm.}^{-2}\text{ sec.}^{-1} \tag{12}$$

This is of the order of 1 per cent. of the total energy radiated by the sun (5.10^{10} erg cm.$^{-2}$ sec.$^{-1}$). At least a considerable fraction of this should be converted into heat in the inner corona.

5.5. Theory of the corona

Since Edlén's identification of the coronal lines (1942) it can be considered as certain that the temperature of the corona is of the order of 10^6 degrees. The problem of how the corona is heated to this exceedingly high temperature has become very important.

The total energy radiated by the corona is estimated by Waldmeier as about 10^{-3} of the solar radiation.

5.51. *Heating of the corona.* The idea that the corona may be heated by some sort of waves, emitted upwards from the photosphere, has been proposed by several authors independently. The waves may be either sound waves (Biermann, 1946; Houtgast, 1947; Schwarzschild, 1948) or magneto-hydrodynamic waves (Alfvén, 1947). Against the former theory it may be objected that in its present form it does not take account of the influence of the solar magnetic field, which certainly affects sound waves very much, especially when the sound velocity is close to the magneto-hydrodynamic wave velocity, which occurs in the solar atmosphere. Against the latter it may be objected that in its present form it does not take account of the compressibility of the medium, which certainly is important in the solar atmosphere. A satisfactory theory may be reached by introducing electromagnetic effects into the sound theory, or—which is the same—by introducing the compressibility of the medium into the magneto-hydrodynamic theory. As in this treatise we attempt to trace electromagnetic phenomena, we shall confine ourselves here to the magneto-hydrodynamic theory of the heating of the corona. The fundamentals of this theory have already been given in the preceding paragraphs.

At first sight it might be astonishing that phenomena in the relatively cool photosphere should be able to cause such a high temperature as is observed in the corona. In reality, however, the process is about the same as that which converts the heat from burning coal in a furnace into the heat of the incandescent filament of an electric lamp. In fact, the heat from the burning coal produces mechanical motion in a steam-engine and in a similar way the temperature gradient in the photosphere produces a mechanical motion, convection. The mechanical motion is converted into electromagnetic energy by a generator in the terrestrial case, and into magneto-hydrodynamic waves (which contain an electromagnetic component) in the solar case. The electromagnetic energy is transmitted by wires, ultimately to, for example, an incandescent lamp, the filament of which may be heated to a temperature which is not at all correlated to the furnace temperature. In a similar way the Joule effect,

when the magneto-hydrodynamic waves are damped, may heat the medium (chromosphere or corona) to a temperature which is determined by the ratio between wave energy and energy loss, but has no direct connexion with the photospheric temperature.

5.52. *The corona as an atmosphere at high temperature.* Immediately after Edlén's identification of the coronal lines it became evident that the corona could be interpreted as an atmosphere at high temperature (Alfvén, 1941 *b*). This idea has been further developed by Waldmeier (1945 *a*).

Most of the light which we receive from the corona is photospheric light scattered by the free electrons in the corona. From photometric observations Baumbach has derived the density, n electrons cm.$^{-3}$, of the electrons:

$$n = 10^8(0{\cdot}036\eta^{-1{\cdot}5}+1{\cdot}55\eta^{-6}+2{\cdot}99\eta^{-16}), \tag{1}$$

where $\eta = R/R_0$.

The works by Grotrian (1934), Öhman (1947), van de Hulst (1947), and Allen (1947) have shown that the corona consists of two components, the K-component or real corona and the F-component which is due to light scattered by dust particles between the sun and the earth. According to a revised formula by van de Hulst the electron density of the real corona

$$n = 10^8(1{\cdot}23\eta^{-6}+2{\cdot}99\eta^{-16}). \tag{2}$$

The negative charge of the electrons must be almost exactly compensated by positive charge from ions (see § 1.4). There is no reason to suppose that the chemical constitution of the corona differs very much from that of other parts of the sun. Hence most of the positive charge in the corona should be due to protons. (Because of the high temperature no neutral hydrogen atoms could be present.) There is also a small number of heavy multiply charged ions (e.g. iron stripped of 14 or 15 electrons) which are responsible for the coronal emission lines.

For a rough estimate of the conditions in the corona, let us assume that it consists of n electrons and n protons per cm.^{3} At least in the inner corona the density is high enough to ensure thermal equilibrium between the 'molecules' (but of course not between molecules and quanta!), so we can apply the common laws of kinetic gas theory. We assume that the mean energy W of the molecules (in our case electrons and protons) amounts to $3kT/2$. As there are $2n$ molecules the gas pressure is given by

$$p = 2nkT. \tag{3}$$

If m_H ($= 1{\cdot}66.10^{-24}$ g.) is the mass of a hydrogen atom and

$$g_\odot \; (= 2{\cdot}74.10^4 \text{ cm. sec.}^{-2})$$

is the acceleration at the sun's surface, the gravitational force acting upon a cubic centimetre is $g_\odot nm_H \eta^{-2}$. As we have assumed that this force is compensated by the pressure gradient, we have

$$\frac{dp}{R_\odot d\eta} = -\frac{g_\odot nm_H}{\eta^2}. \tag{4}$$

Differentiating (3) we obtain from (3) and (4)

$$\frac{d}{d\eta}\left(\frac{T}{T_0}\right) + \frac{1}{n}\frac{dn}{d\eta}\frac{T}{T_0} = -\frac{1}{\eta^2}, \tag{5}$$

where $$T_0 = \frac{g_\odot R_\odot m_H}{2k} = 11{\cdot}6.10^{6\circ}\,\mathrm{K}.$$

From (5) we obtain $$\frac{T}{T_0} = -\frac{1}{n}\int \frac{n}{\eta^2}\,d\eta, \tag{6}$$

or, according to (2),

$$\frac{T}{T_0} = \frac{(1{\cdot}23/7)\eta^{-7} + (2{\cdot}99/17)\eta^{-17}}{1{\cdot}23\eta^{-6} + 2{\cdot}99\eta^{-16}}. \tag{7}$$

The value of T from this formula is shown in Fig. 5.2. The temperature in the inner corona is almost constant, having the value of about 1,000,000°.

Taking account of the fact that the corona contains heavier gases than hydrogen, we obtain a still higher value for the temperature.

5.53. *Fine structure of the corona.* As has been shown above, the corona could be interpreted as an atmosphere, heated to about 1,000,000° by electric currents associated with the granulation waves. This result refers to the average conditions, which, however, are frequently disturbed.

First it should be observed that, owing to the magnetic field, the corona is highly anisotropic. The electric conductivity parallel to the magnetic field is much greater than perpendicular to the field. Magnetohydrodynamic waves move parallel to the field. Charged particles move easily along the magnetic lines of force, but a motion perpendicular to the field is braked by electromagnetic phenomena.

If we suppose that in the photosphere the generation of granulation waves is uniform over the whole solar surface, it is easily seen from the geometry of the magnetic field that in the corona the equatorial region must receive more energy than the polar regions. Hence we should expect the temperature to be higher at low latitudes than at high latitudes. This is in agreement with the observational fact that the corona is normally more extended over the equator than over the poles.

Moreover, as Waldmeier (1945 *b*) has shown, the ionization in the corona is higher at low latitudes. Whereas the red corona line 6374, which according to Edlén is emitted by Fe X which has an ionization energy of 233 e.v., is observed in all latitudes, the green line 5303 deriving from Fe XIV with an ionization energy of 355 e.v. is essentially confined to low and middle latitudes but in general absent near the poles.

Even if we assume the granulation waves to supply the 'normal' heating of the corona, there is no doubt that other heating processes are also active. Of special importance are effects deriving from the prominences. As we shall see in § 5.6, prominences may be explained as electrical discharges in the solar atmosphere. Electric currents in the prominences, and maybe also secondary current systems in their environment, heat the corona in excess of the heating supplied by the electric currents of the granulation waves. The dependence of the coronal structure on the solar activity may be explained in this way. In fact a prominence or any other sign of solar 'activity' usually affects the corona in such a way that it becomes especially brilliant or extended over the active region. This phenomenon is often supposed to be due to an 'ejection' of matter from the photosphere, but the mechanism of ejection is left unspecified. It seems more natural to assume that the heating of the corona—or in part also of lower layers—is the primary effect, resulting in a thermal expansion of the corona or (as will be discussed in § 5.7) a diamagnetic expulsion of the heated gas. This is also in agreement with the interesting result by Waldmeier (1945 *b*) that the yellow corona line 5694 is observed almost only above sunspot groups of great activity. This line derives from Ca XV, which has an ionization energy of as much as 814 e.v., and can be considered as an indicator of extreme temperatures.

5.6. Prominences

Prominences are most easily observed when situated at the solar limb. Some characteristic types are seen in Figs. 5.23–5.25. At the disk they are visible as long filaments (see Fig. 5.25). Two different types can be discerned: *Eruptive prominences* are usually associated with sunspots and solar flares. They change their structure very rapidly and are shortlived (hours or days). *Quiescent prominences* are more persistent and may last for months. They may be as long as 40 or 50 heliographic degrees. According to d'Azambuja (1941) their average dimensions are: length 6.10^{10} cm., height 9.10^{9} cm., breadth 3.10^{8} cm.

In the prominences internal motions with velocities of the order

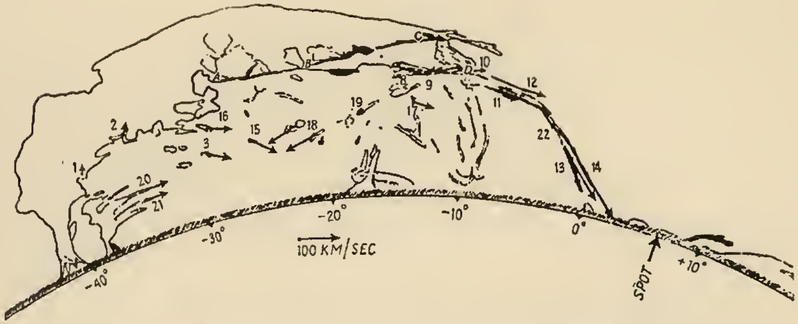

Fig. 5.23. Splash prominence (Pettit, *Publ. Yerkes Obs.* 3, 221, 1925).

Fig. 5.24. Prominence with associated sunspot group (Öhman).

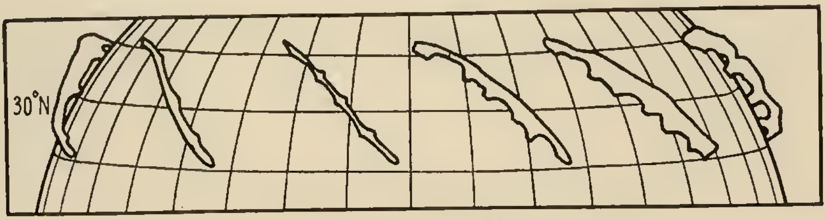

Fig. 5.25. Quiescent prominence (d'Azambuja).

10^6–10^8 cm. sec.$^{-1}$ are observed. The accelerations are frequently much larger than could be produced by gravitation, and their directions make arbitrary angles with the vertical. Matter is often ejected upwards and sideways, but the most characteristic motions in prominences are directed downwards, frequently towards a sunspot, which seems to 'attract' the prominence. In some of the marvellous moving pictures by Lyot and Menzel matter is seen to move at almost constant speed along a fixed path which could very well be a magnetic line of force of a sunspot field. Matter seems to be brought into the prominence high up in the corona and to stream downwards to the photosphere.

The spectrum of prominences is similar to that of the chromosphere. The excitations correspond to a temperature of the order 10,000–20,000°. (See Waldmeier, 1941, p. 225.)

Many theories have been proposed in order to explain the prominences. They have been supposed to constitute an analogy to terrestrial volcanoes. This is an early hypothesis and cannot be taken seriously with our present knowledge of solar constitution, but still different authors sometimes assume that 'hot gases are ejected' from the sun's interior, where the temperature is high. Such a process is of course absolutely impossible. The hot gas in the sun's interior is at an enormous pressure and if this pressure is reduced the gas will cool adiabatically. If matter really were brought up from the sun's interior into the photosphere, it would be much cooler than the photosphere. (Compare the upward motion in sunspots.) No attempt has ever been made to formulate the hydrodynamic conditions for the supposed 'ejection' of matter from the photosphere upwards. Moreover, in reality, matter usually streams downwards in the prominences.

The radiation pressure theory of prominences is not very much more successful. Certainly if we consider a small part of a prominence, which is accelerated, it is in principle possible to suppose that the acceleration takes place under the action of a force composed of gravitation and radiation pressure. In some cases the postulated radiation pressure must be very large, requiring extremely high temperatures, but very high temperatures are required anyhow, as indicated by observations of solar flares and radio noise. The real difficulties start when we try to explain the motions in the whole prominence, or, still worse, a system of coexisting prominences. Take for example a quiescent prominence, which may be an arc extending over more than 50 heliographic degrees. In this arc matter moves, upon the whole horizontally, and at a rather uniform speed. How should the light sources, responsible for the

radiation pressure, be situated in order to produce such a type of motion? And why does the radiation pressure affect matter only in that very thin arc which constitutes the prominence, and not disturb the conditions in the photosphere around it?

It would be still more difficult to explain the prominences visible in some of the moving pictures. These may form a complicated pattern of fountains, with matter moving in different directions but usually along certain fixed paths.

When studying prominences it is difficult to avoid the impression that some force other than gravitation and radiation pressure is the really important factor. It may be worth while to investigate whether this unknown force is of electromagnetic origin. Hence we ought to study the electromagnetic conditions in the solar atmosphere.

First of all it is of interest to point out that there is an important difference between the terrestrial and the solar atmospheres. The terrestrial atmosphere is an electric insulator, whereas the solar atmosphere is a good conductor (see § 5.21). Hence so-called electrostatic phenomena cannot exist in the solar atmosphere. In the terrestrial atmosphere a thundercloud may keep a charge at many million volts during a long time. In the solar atmosphere the charge is carried away long before it has been able to produce a notable potential difference.

Consequently there is no similarity in electrical respects between the solar atmosphere and the lower part of the terrestrial atmosphere [as supposed by Bruce (1946)]. It is more legitimate to make a comparison with the highest part of our atmosphere, the ionosphere, which is an electrical conductor, although not so good as the solar atmosphere.

5.61. *Discharge theory of prominences. Electromotive force.* We shall here discuss a theory of solar prominences according to which they constitute electric discharges. The electromotive force is supposed to be generated by induction which occurs when solar matter moves in the sun's general field and the sunspot magnetic fields. These fields also affect the discharge itself, because in a magnetic field the conductivity becomes highly anisotropic, so that the current prefers the direction of the magnetic field. As we have seen in § 3.21, an electric field with a component $E_{\parallel}$ parallel to the magnetic field produces a current $i_{\parallel}$ which equals $\sigma_{\parallel} E_{\parallel}$, whereas an electric field $E_{\perp}$ perpendicular to the magnetic field in the first moment gives a current $i_{\perp}$ equal to $\sigma_{\perp} E_{\perp}$. After some time this current will have decayed because the medium has been accelerated and an opposing electromotive force produced. Hence,

roughly speaking, the current can flow only along the magnetic lines of force. *The condition for a discharge above the solar surface is that a magnetic line of force intersects the surface in two points of different potential.*

We have seen in § 5.37 that sunspots are associated with very powerful electromotive forces. A sunspot-producing magneto-hydrodynamic whirl is connected with an induced e.m.f. of the order of 10^8–10^9 volts, which, however, under ideal conditions (homogeneous magnetic field, constant mass density, infinite conductivity) is exactly compensated by a polarization e.m.f. If the ideal conditions are not satisfied we should not expect the compensation to be exact. In the case treated in § 5.37 the compensation is exact below the surface but the induced e.m.f. is uncompensated above the surface. Even if only a relatively small part of the voltage is left uncompensated, a very large e.m.f. is produced.

There is also another process which may produce very high voltages. Sunspots are associated with an upward motion in the spot itself and an outward motion in the environment (Evershed effect). The outward motion is acted upon by the Coriolis force, which produces a tangential component and gives a vortical motion around the spot. Such motions have actually been observed by Fox (1921). Of course they give rise to magneto-hydrodynamic waves which are moving downwards.

Corresponding to the outward motion in the upper layer there is an inward motion in the deeper layer which also gives rise to a tangential motion in the opposite direction to that mentioned above. This is transmitted upwards by magneto-hydrodynamic waves and may reach the surface. Hence the surface motion may go in either direction depending upon which effect preponderates.

Together with the magnetic field the tangential motion gives a polarization e.m.f. in a radial direction. The voltage

$$V = \frac{1}{c}\int [\mathbf{vH}]\,\mathbf{ds}. \tag{1}$$

This voltage may also be considered as due to the electric field associated with the magneto-hydrodynamic wave which transmits the tangential motion. In a region with swiftly varying density (such as in the photosphere) it may be largely uncompensated.

Tangential displacements of 0·05 heliographic degree/day are frequently observed (Fox, loc. cit.). This corresponds to a velocity of 10^3 cm./sec. With a magnetic field of 200 gauss, which is reasonable

near a sunspot, this gives an electric polarization of 2.10^{-3} volts/cm. When integrated over a distance of a few heliographic degrees (say, 5.10^{9} cm.) a voltage of 10^{7} volts is obtained. Hence also in this way a very high e.m.f. is produced.

The mechanisms which we have discussed are dependent upon the conditions around sunspots which are rapidly changing, so that the fields produced in this way may not in general be very long-lived. There

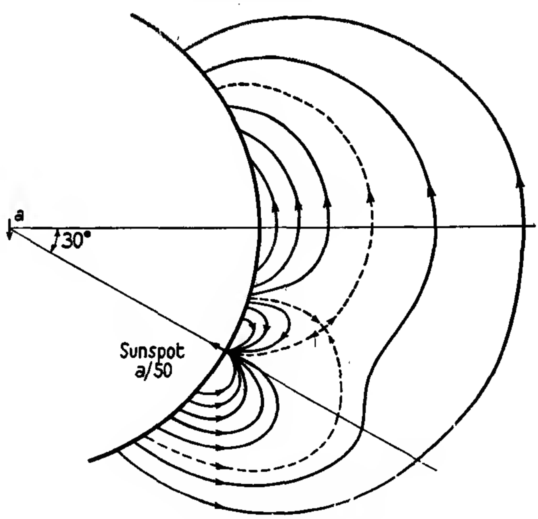

FIG. 5.26. Magnetic field disturbed by sunspot field.

may also be a possibility that longer lasting voltage differences are produced in the following way.

Due to the non-uniform rotation, the sun is polarized even if seen from a coordinate system which takes part in the rotation (compare §§ 1.3 and 5.82). As the general magnetic field is symmetrical with respect to the axis of rotation, which according to § 5.24 is necessary, each surface generated by rotating a magnetic line of force is an electric equipotential surface. Hence in the undisturbed case the electric field is always perpendicular to the magnetic field and no discharge is possible.

If the magnetic field is disturbed, e.g. by a sunspot field, a magnetic line of force may intersect the solar surface at two points of different voltage. Hence a discharge may be produced as we have found above. In this case the e.m.f. is produced by the non-uniform rotation of the sun, which means that it is much more constant than if produced by some local motion.

The polarization produced by the non-uniform rotation and referred to a system in which the equator is at rest may be computed from

$$V = \frac{1}{c}\int\limits_0^{\varphi} [vH]R_0\,d\varphi, \tag{2}$$

where R_0 is the solar radius, φ the latitude, H the magnetic field, and v the velocity

$$v = R_0\cos\varphi(\Omega_\varphi - \Omega_e). \tag{3}$$

According to 5.24 (20) the angular velocity $\Omega_\varphi = \Omega_e - \Omega'\sin^2\varphi$.

Thus we obtain

$$V = V_0\sin^4\varphi, \tag{4}$$

where $V_0 = \frac{1}{4}c^{-1}\Omega' R_0^2 H_p = 0{\cdot}58.10^6$ e.s.u. $= 1{\cdot}7.10^8$ volts.

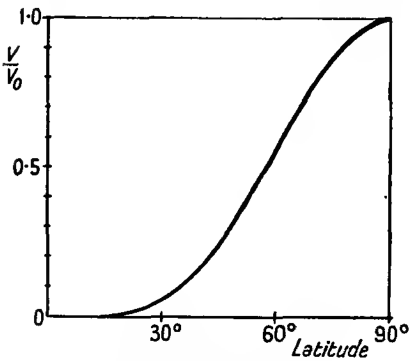

Fig. 5.27. Polarization due to non-uniform rotation.

This function is plotted in Fig. 5.27. If a magnetic line of force from a sunspot at the equator ends at, for example, $\varphi = 50°$, a voltage difference of $0{\cdot}6.10^8$ volts is obtained. To this should be added the voltage due to polarization in the spot field. Of course V is also somewhat changed by the change in magnetic field produced by the spot.

The electromotive force associated with the non-uniform rotation could be expected to be more constant and long-lived than the voltages from local motions near sunspots, because much bigger masses take part in the former motion.

5.62. *Structure of prominences.* We have seen that extremely powerful electromotive forces, maybe of the order of 10^7–10^8 volts, are likely to be active in the solar atmosphere. It is quite natural that these should produce discharges. As above the lower part of the chromosphere the conductivity is much larger parallel to a magnetic field than perpendicular to it, we should expect discharges above the solar surface to follow the magnetic lines of force. Frequently the visibility of different parts of the discharge may differ very much with the result that often only a part of it is observed. An example of a fully developed discharge is given by the 'splash prominence' of Fig. 5.23. In terms of the theory it should be interpreted as a discharge following magnetic lines of force which go from a big spot near the equator and intersect the solar surface at $-40°$ latitude. A small discharge occurs to the right of the spot, and also some secondary discharges are visible below the main discharge.

In many cases only that part which is remotest from the spot is visible at first, often as a rather diffuse cloud, which later sends out a 'streamer' towards the spot. The characteristic phenomenon that spots 'attract' prominences may be due to the fact that most lines of force end in spots, and that, for some unknown reason, the visibility of the discharge is initially low near the spot.

Especially near large spot groups the magnetic lines of force often form a very complicated pattern, in which discharges cause a system of 'fountain prominences'.

It must be observed that the magnetic lines of force are not rigid. On the contrary, the discharge may change their shape, and frequently one gets the impression that they are pressed down under the load of the matter in the prominence. This occurs especially in long quiescent prominences, which are almost horizontal (see Fig. 5.25). Often the discharge is branched so that it looks like a series of arcs.

What is said indicates that prominences may be identified with electric discharges.

Considering the problem more closely, several objections arise. The shape of the prominence shows that it must be a constricted discharge. In an ordinary constricted gas discharge the discharge channel has a higher conductivity, a higher temperature, and usually also a lower density than the surroundings. In an arc at atmospheric pressure, for example, the temperature is more than one power of 10 above that of the surroundings, and the density correspondingly lower. The discharge channel is a good conductor, whereas the surrounding gas is an insulator. In the solar case the conditions are reversed. The corona is normally very hot, about 10^6 degrees, and the prominences are certainly much cooler, probably 10,000–30,000°. The density in the prominence is at least 100 times that of the corona. As the corona is normally completely ionized a discharge in it could not increase the ionization appreciably. (The relative number of heavy atoms which could give off more electrons is small.) According to formula 3.23 (64) the conductivity of a completely ionized gas is proportional to $T^{\frac{3}{2}}$ and independent of pressure. Hence the conductivity in a prominence must be lower than in the surroundings, and it is difficult to understand why the discharge current should flow only in the prominence. Further, the internal motions in a prominence are usually directed downwards. In a prominence arc above the solar surface motions may go downwards in both branches. If the observed motions are assumed to be due to the motion of positive ions, these must go in the same direction in the whole prominence, so

that if they are directed downwards in one branch they must go upwards in the other.

All these objections are valid only if we suppose that a discharge through a completely ionized gas of cosmic dimensions, such as the corona, must have essentially the same properties as a discharge through a cold gas in the laboratory. The analysis in § 3.41, although rudimentary, indicates that this is not at all the case.

A high current discharge in an ionized gas is according to § 3.41 likely to constrict because of the electromagnetic attraction between parallel currents. As soon as a current has started a magnetic field is produced, the lines of force (in the absence of a primary magnetic field) encircling the current. The path (in the centre of the current) where the magnetic field is zero is of special importance and will here be called the axis. In the whole space occupied by the current, matter is accelerated towards the axis until the pressure at the axis is strong enough to compensate the mutual electromagnetic attraction of the currents. The pressure may be calculated from formula 3.4 (5). The difference, Δp, between the pressure at the axis and at infinity is of the order of $1/4\pi\, H_m^2$, where H_m means the maximum magnetic field produced by the discharge. Putting this as low as 10 gauss, we find $\Delta p = 10$ dyn. cm.$^{-2}$ This is very great in comparison with the gas pressure in the corona, which for a temperature $T = 10^6$ degrees and a density of $n = 10^8$ molecules cm.$^{-3}$ is given by $p = nkT = 1{\cdot}4 . 10^{-2}$ dyn. cm.$^{-2}$

We have found that a discharge through the corona brings matter from the surroundings towards the discharge axis. The matter thus concentrated can escape only along the axis, and will in general stream down to the photosphere. This explains the predominant type of motion in a prominence.

The low temperature of the prominence in comparison with the corona may simply be an effect of the increased pressure. The corona can only be kept at a high temperature because its density is extremely low. An increase in density means an increase in radiation losses. Certainly a Joule heating takes place in the prominences, but this does not suffice to cover the increased losses through radiation. Joule heating by stray currents in the environment of the real prominence may heat the corona, so that the coronal temperature near a prominence may rise in spite of the fact that the prominence is rather cool.

If we try to estimate the currents in a prominence a quantitative difficulty seems to arise. The magnetic field of the discharge current cannot be as strong as 1,000 gauss, because such a field should have been

detected by the Zeeman effect. Suppose that a prominence is a thin sheet, thickness $x = 3.10^8$ cm. (compare § 5.6), within which the current density i is constant. Then for the magnetic field at the surface

$$H = \frac{2\pi}{c}ix. \tag{5}$$

If $H = 100$ gauss, we find

$$i = 1{\cdot}6.10^3 \text{ e.s.u.}$$

With $\sigma = 10^{13}$ e.s.u., we obtain

$$E = i/\sigma = 1{\cdot}6.10^{-10} \text{ e.s.u.}$$

Integrated over the length 6.10^{10} cm. of the prominence, this gives a voltage difference of 10 e.s.u. = 3,000 volts, which is very much less than expected (10^7–10^8 volts).

This discrepancy is due to the fact that the current is not determined by the resistance but by *the inductance*.

Consider the circuit consisting of a prominence and the solar surface. Denote its resistance by X, and its inductance by L. An e.m.f. V produces a current i. We have

$$V = Xi + L\frac{di}{dt}. \tag{6}$$

As an approximation let us regard the circuit as a circular loop (radius R) with circular cross-section of radius ρ. If the conductivity is σ we have according to well-known formulae

$$X = \frac{2R}{\sigma\rho^2}; \tag{7}$$

$$L = \frac{4\pi R}{c^2}\left(\ln\frac{R}{\rho} + \frac{1}{3}\right). \tag{8}$$

The time-constant τ of the circuit is given by

$$\tau = L/X = \frac{2\pi\sigma}{c^2}\rho^2\left(\ln\frac{R}{\rho} + \frac{1}{3}\right). \tag{9}$$

Putting $\rho = 3.10^8$ cm., $R = 10^{10}$ cm., $\sigma = 10^{13}$ e.s.u. (all moderate values), we obtain

$$X = 2.10^{-20} \text{ e.s.u.} = 2.10^{-8} \text{ ohm,}$$
$$L = 5.10^{-10} \text{ e.s.u.} = 500 \text{ henry,}$$
$$\tau = 2{\cdot}5.10^{10} \text{ sec.} = 800 \text{ years.}$$

The big value of τ indicates the enormous importance of the inductance. For times smaller than τ the current in a circuit is determined mainly by the inductance. Hence it is only when V does not change

during several centuries that the current system is determined by the conductivity. More rapidly changing voltages give currents mainly determined by the inductance.

If a constant voltage has been active during the time t ($\ll \tau$) we have approximately

$$\frac{Xi}{V} = \frac{t}{\tau}. \tag{10}$$

If, as in the example discussed above, the electromotive force $V = 3.10^7$ volts and the voltage drop over the resistance $Xi = 3{,}000$ volts, we have $t = 3{,}000\tau/3.10^7 = 10^{-4}\tau = 2{\cdot}5.10^6$ sec. Hence the given figures would be characteristic for a quiescent prominence which is about one month old.

The analogy with an ordinary electric circuit is illuminating in some respects, but it must not be taken too literally. In the solar atmosphere there is no fixed circuit because the magnetic lines of force are flexible and, moreover, influenced by the current which produces a secondary magnetic field. The most important difference is, however, that the above calculation refers to a current in a wire surrounded by an insulator, whereas in the solar atmosphere the surrounding medium is a conductor. Hence we come close to the well-known 'skin effect' problem: If an electric field is applied to a solid conductor, the current starts near the surface. At first no current flows in the interior because induction produces an e.m.f. which compensates the applied voltage. The surface current penetrates more and more deeply into the conductor, so that after some time a current flows in the whole of it.

As shown in § 3.41, the 'skin effect' in a compressible conductor is probably very different from the common effect which is characteristic for a solid conductor. The current does not flow at the surface but in a channel in the interior of the compressible medium. Induction protects the rest of the medium from currents.

The theory of prominences discussed here is of course only a first approach to the problem. Before a real theory could be developed, the 'inverse skin effect' must be studied closely. The effect of an imposed magnetic field parallel to the current must be investigated.

5.7. Emission of ion clouds

Magnetic disturbances, including aurorae, occur frequently on the earth about one day after a disturbance in the sun (solar flare) or after the meridian passage of a big sunspot. The conclusion has been drawn that some agent is emitted from the sun, which travels at such a speed

($\approx 2.10^8$ cm. sec.$^{-1}$) that it reaches the earth after about one day. This agent cannot be an electron or ion beam containing particles of one sign only, because as Schuster has pointed out, such a beam cannot transmit power enough to produce a magnetic storm. Instead it must be electrically almost neutral, so that it contains about the same amount of positive and negative particles. A theory on this basis has been developed by Chapman and Ferraro (1929), later leading to a theory of magnetic storms (see § 6.1).

The most difficult question is how the ion cloud could be accelerated to such a high velocity. A proton with the velocity 2.10^8 cm. sec.$^{-1}$ has a kinetic energy of more than 20,000 electron volts, so that even if the cloud consists of protons and electrons only, the average energy per particle must exceed 10,000 e.v. This corresponds, with random distribution, to a temperature of 10^8 degrees.

If possible one would avoid assuming the existence in this connexion of such high energies. Certainly the recurrent appearance of disturbances one day after the meridian passage of a sunspot is no definite argument for such high velocities, because a beam sent out from the sun may take part in the solar rotation and for some reason be situated in such a way that it passes the earth after a delay of one day. The occurrence of a storm one day after a solar flare constitutes a much stronger argument. Something must have been emitted which reaches the earth one day later. This need not be an ion cloud but could be, for example, some sort of waves. A decisive argument for the ion cloud hypothesis seems to be the fact that in the solar atmosphere expulsions of prominences are observed with velocities up to 10^8 cm. sec.$^{-1}$ This shows that some mechanism in connexion with the prominences is able to impart such high velocities to a cloud.

An ionized gas which is heated to a high temperature is strongly diamagnetic in the absence of limiting walls (see § 3.3). When situated in an unhomogeneous magnetic field it is acted upon by a force which tries to bring it out of the field. If it is free to move under the action of this force it cools down, and converts its kinetic energy into velocity in the direction of the force. In this way an ion cloud may be expelled at any velocity, provided only that it is initially heated to a sufficiently high temperature. For a velocity of 2.10^8 cm. sec.$^{-1}$ this temperature is, as stated above, of the order of 10^8 degrees. It should be observed that it suffices if the electronic temperature reaches this value, so the gas need not be so hot. Anyhow this is certainly a very high temperature and we must hesitate to assume that some part of the solar atmosphere

could be so intensely heated. On the other hand, we know that the corona is quite normally at a temperature of 10^6 degrees, and in a discharge an electronic temperature 100 times the gas temperature is not at all remarkable.

A temperature of 10^8 degrees corresponds to particle energies of the order of 10^4 e.v. The electromotive force of a prominence discharge has been estimated to be of the order of 10^8 volts (see § 5.61). Hence the voltage is no doubt sufficient to give the charged particles an energy as high as required. The main difficulty, however, is how it could be so concentrated as to give rise to such high temperatures. We have seen that most of the primary voltage is compensated by inductive effects and that only a few thousand volts are available in the discharge (see § 5.62).

In § 3.5 it has been shown that in an ionized gas a discharge is not stable when the current density exceeds a certain limit. This instability is due to the fact that, when the electron temperature increases, the interaction between the electron gas and the ionic gas decreases, with the result that the cooling of the electronic gas becomes less effective, so that the temperature increases still more. The result is likely to be an 'electron gas explosion', leading to an acceleration of electrons to very high velocities. When the current in a circuit is interrupted, most of the electromagnetic energy of the circuit is dissipated at the place of interruption, where the voltage difference may be much higher than the initial e.m.f. (because to this is added an induced voltage). Hence it seems reasonable that very high temperatures may be reached by the electronic gas.

5.8. Electromagnetic conditions around the sun

There is almost no observational evidence concerning the important question of the electric and magnetic conditions in the environment of the sun, say within the planetary system. Especially the problem of the electric field seems never to have been attacked seriously from the theoretical or the observational side.

5.81. *Magnetic fields.* According to § 5.22 the sun is likely to possess a magnetic field which at the surface and outside is approximately a dipole field. In the neighbourhood of the earth the terrestrial magnetic field is superimposed on this field. At a distance of $4 \cdot 10^{10}$ cm., i.e. the order of magnitude of the distance to the moon, the solar and terrestrial fields are equal, each of them amounting to 10^{-6} gauss. We know nothing about the magnetic fields of other planets, but if magnetic moment per

unit volume is of about the same order as in the case of the earth, even these fields are confined to the close neighbourhood of the planets.

Outside the solar magnetic field there may be a general galactic magnetic field (see § 7.5). Arguments from the properties of cosmic radiation indicate that the galactic field may have a strength of between 10^{-12} and 10^{-9} gauss. The solar magnetic field has decreased to 10^{-10} gauss at a distance of $0{\cdot}3 . 10^{15}$ cm. (assuming a dipole field), which is near Neptune's orbit.

Consequently the magnetic conditions within the whole planetary system, except in the close vicinity of the planets, is likely to be governed by the solar magnetic field.

Currents in interplanetary matter may disturb the solar field. We do not know to what extent this takes place.

5.82. *Electric fields.* The electrical conditions are not so easy to analyse.

Suppose that an electrically conducting, uniformly magnetized sphere revolves around the axis of magnetization with the constant angular velocity Ω. Suppose further that in a reference system which follows the rotation, no electric fields exist. Seen from a fixed reference system an electric field E is induced by the motion with the velocity v in the magnetic field H:

$$\mathbf{E} = \frac{\mu}{c}[\mathbf{vH}]. \tag{1}$$

Consequently the voltage at a certain latitude φ_0 of the sphere is

$$V = \frac{1}{c}\int H_v \Omega R_0 \cos\varphi_0 \, R_0 \, d\varphi_0, \tag{2}$$

where R_0 = the radius of the sphere and H_v ($= H_p \sin\varphi_0$) is the vertical component of the magnetic field. We obtain

$$V = -\frac{H_p \Omega}{2c} R_0^2 \cos^2\varphi_0 + V_p, \tag{3}$$

where H_p and V_p are the magnetic field and the electric potential at the pole. For the sun we obtain for the voltage difference between the equator and one of the poles, putting $H_p = 25$ gauss,

$$V_e - V_p = 1{\cdot}7 . 10^9 \text{ volts.} \tag{4}$$

For the earth we obtain ($H_p = 0{\cdot}6$ gauss)

$$V_e - V_p = 0{\cdot}88 . 10^5 \text{ volts.} \tag{5}$$

From (3) we know the potential distribution at the surface of the sun. If we calculate the conductivity of interplanetary space from the

formulae of Chapter III, it would be possible to find the current system produced by this potential, and hence we could also compute the potential distribution in interplanetary space.

In reality the problem is more complicated, because, as in § 5.62, the current is not determined by the conductivity but by the inductance, and as we deal with much larger linear dimensions the relative importance of the inductance is much greater. Any change in the current system will produce induced electromotive forces whieh tend to oppose the change. Hence if an electric field is produced, e.g. by a space charge, this field is compensated by an induced field. As the induced fields are not derivable from a potential *it is meaningless to speak of an electrostatic potential in the environment of the sun.*

The current density i may be computed from

$$\mathbf{i} = \sigma\left\{\mathbf{E}+\frac{\mu}{c}[\mathbf{vH}]\right\}. \tag{6}$$

If there is a density gradient the diffusion current should be added (see § 3.25), but this is of little importance for the following. Further, a Hall current is produced if $\omega\tau$ is large. If the current is extended over a volume with the linear extension x, the magnetic field from i is of the order

$$H \approx \frac{1}{c}xi. \tag{7}$$

In the environment of the sun, the linear dimensions are at least of the order $x = 10^{11}$ cm. Except very close to the sun we should not expect stationary magnetic fields as strong as, say, 10 gauss. Hence we find

$$i < \frac{Hc}{x} = \frac{10\cdot 3.10^{10}}{10^{11}} = 3 \text{ e.s.u.} \tag{8}$$

Putting $\sigma = 10^{13}$ e.s.u., we obtain

$$\left|\mathbf{E}+\frac{\mu}{c}[\mathbf{vH}]\right| < 3.10^{-13} \text{ e.s.u.} = 10^{-10} \text{ volt/cm.} \tag{9}$$

Compared with other fields, e.g. the field deriving from the polarization of the sun, this is negligible.

The conclusion is that as an approximation we could put

$$\mathbf{E} = -\frac{\mu}{c}[\mathbf{vH}]. \tag{10}$$

The electric field at a certain point is (approximately) determined by the magnetic field and the state of motion at the point in question. The electric field is not usually derivable from an electrostatic potential.

This result may be compared with the example given in § 5.62. Of an applied voltage of 10^7–10^8 volts most is compensated by induced fields and only 3,000 volts are left to produce a current. Because of the larger dimensions the conditions in the environment of the sun may be expected to be still more extreme.

If a stream of matter with velocity v is supposed to be part of a magneto-hydrodynamic wave, equation (10) is exact and no phenomena at all should occur outside the stream (see § 4.6). In a magneto-hydrodynamic wave there is induced an electric field E which exactly compensates the polarization. According to formula 4.6 (5) the induced field is the same as that given by (10). In the ideal case (homogeneous magnetic field, incompressible liquid with constant density) no disturbance occurs outside the moving matter. As any stream in a magnetic field could be considered as composed of magneto-hydrodynamic waves equation (10) should hold exactly under ideal conditions, and as a fair approximation under conditions not too remote from these.

Close to the sun, say within the region of the corona, it is likely that the average field, seen from a *rotating* system, is small (non-uniform rotation neglected), which means that the matter rotates with about the same angular velocity as the sun. Far away from the sun the average field seen from a *fixed* system is probably small and the matter at rest. We do not know where the limit between the rotating and resting matter is located. Probably the state of motion of interplanetary matter changes very much, especially in connexion with magnetic storms.

5.9. Solar noise

If the aerial system of a very sensitive short-wave radio receiver is directed towards certain parts of the sky the normal noise of the apparatus increases. In this way Jansky (1933) showed that the Milky Way emits radio waves. Later Southworth (1945) and Reber (1944) found that the sun also is a radio transmitter. As Appleton (1945) and Hey (1946) have shown, the intensity of the signals increases greatly during periods of solar activity. According to interference measurements by Ryle and Vonberg (1946) and McCready, Pawsey, and Payne-Scott (1947), the transmitting area is much smaller than the solar diameter and may not be larger than a sunspot.

Solar noise has been received on a frequency ranging from about 15 to 10,000 megacycles. The lower limit is governed by the opacity of the ionosphere, the upper limit by technical difficulties in constructing sensitive receivers. The signals are heard as a noise which means that

the amplitude varies at random in the same way as in thermal radiation. As therefore the waves may be supposed to be generated by thermal motion of electrons, we can give the 'equivalent temperature' T of the transmitter as a measure of the intensity. In fact for the low-frequency part of the spectrum, Planck's radiation formula can be reduced to

$$S\,d\nu = \frac{2k}{c^3}\,\Omega T\nu^2\,d\nu,$$

where $S\,d\nu$ is the energy flux in the frequency band $d\nu$, k Boltzmann's constant, c the velocity of light, ν the frequency, and Ω the solid angle of the source.

At low solar activity the intensity S at $\nu = 3{,}000\text{–}10{,}000.10^6$ sec.$^{-1}$ (= 3,000–10,000 Mc., i.e. 10–3 cm. wave-length) corresponds to $T = 20{,}000°$ K. supposing the whole solar disk to radiate.

At $\nu = 200.10^6$ sec.$^{-1}$ (1·5 metres) the equivalent temperature is 500,000° K., but increases sometimes to 10^8 degrees when active sunspots and especially solar flares occur (McCready, Pawsey, Payne-Scott, 1947). Taking account of the fact that the source is much smaller than the solar disk, we reach values of a few times 10^9 degrees (cf. also Ryle and Vonberg, 1946).

The disturbances on frequencies of the order 10^8 sec.$^{-1}$ (100 Mc.) consist frequently of sudden large increases in intensity of a duration of some seconds or minutes. They are usually coincident with solar flares and radio fade-outs but, as Payne-Scott, Yabsley, and Bolton (1947) have demonstrated, there is a certain time delay between the disturbances at different frequencies, so that the disturbance at 200 Mc. comes a few minutes earlier than the disturbance at 100 or 60 Mc.

As found by Martyn (1946 *a*), the waves from sunspots are circularly polarized.

Attempts to explain the radio wave emission theoretically have been made by Martyn (1946 *b*), Shklowsky (1947), Kiepenheuer (1946), Giovanelli (1948), and others. The quiet-sun radiation may not be so difficult to account for. Waves of different frequencies originate at different heights in the solar atmosphere. Applying ionospheric formulae Martyn (1946 *b*) has shown that the corona is opaque for 100 Mc. but transparent for 1,000 Mc. Hence the former frequency emanates from the corona and its intensity should correspond to 10^6 degrees, whereas the higher frequencies derive from lower and cooler layers (chromosphere).

The very high equivalent temperatures ($> 10^9$ degrees) observed

during solar flares are more difficult to account for. Giovanelli assumes that in the solar atmosphere there really are electron temperatures of this order (corresponding to more than 10^5 e.v. per electron) and accounts for the production of these through a mechanism somewhat similar to that discussed in § 5.6. As long as possible we should hesitate to assume such high temperatures, but, as seen in § 5.7, the energy per particle in an emitted ion cloud is only one order of magnitude less than this. Martyn and Shklowsky assume instead that plasma oscillations in the

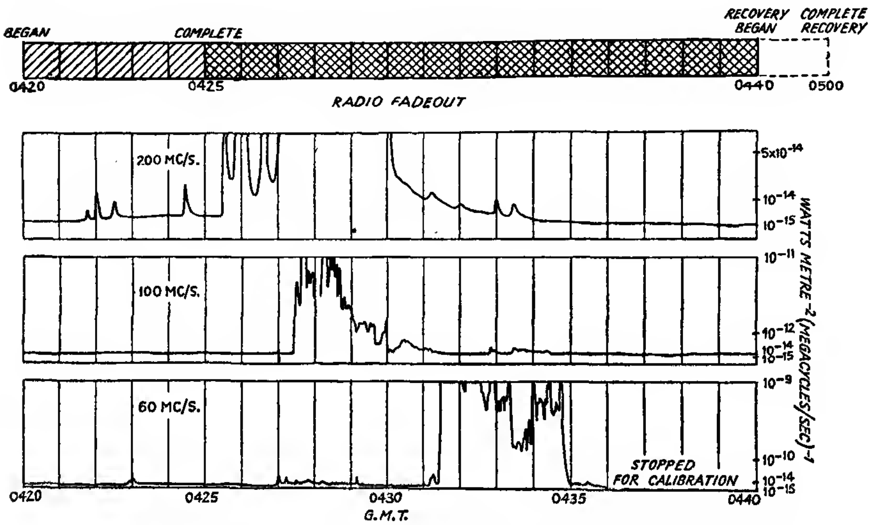

FIG. 5.28. Solar noise. Violent increase first noted on 200 Mc., later on 100 Mc., and still later on 60 Mc. (After Payne-Scott, Yabsley, and Bolton.)

corona are responsible for the emission. Although this idea has not been worked out in detail, it is very interesting. It is known that in the presence of a magnetic field gaseous discharges produce 'noise' oscillations. Even if no gas is present, electrons moving in a magnetic field produce much more noise than expected theoretically, as shown by measurements by Åström on trochotrons (1948) (compare § 2.7). Even if at present we are far from understanding the emission of radio waves from the sun, it can be considered as likely that they must be produced directly, or indirectly, by electric discharges. Hence the discovery of solar noise has stressed the importance of studying electrical phenomena in the sun.

The origin of the galactic noise is still more uncertain. If it is produced in the same way as solar noise, and hence derives from the stars, the

average output of a star must be much larger than that of the sun, because we receive on an average about the same noise from the galaxy as from the sun. This would mean that electric phenomena should be much more prominent on other stars than on the sun. On the other hand, it is also possible that the noise derives from interstellar space.

REFERENCES

General references:

WALDMEIER, M.: *Ergebnisse und Probleme der Sonnenforschung*, Leipzig, 1941.

NICOLET, M.: *Introduction à l'étude des relations entre les phénomènes solaires et terrestres: le Soleil*: Inst. Royal météor. de Belgique, **11**, 1943.

ALFVÉN, H., 1940: 'Tentative Theory of Solar Prominences', *Ark. f. mat., astr. o. fysik*, **27 A**, No. 20.

—— 1941 *a*: 'On the Solar Corona', ibid. No. 25.

—— 1941 *b*: 'Remarks on the Rotation of a Magnetized Sphere, with Application to the Solar Rotation', ibid. **28 A**, No. 6.

—— 1943 *a*: 'On the Effect of a Vertical Magnetic Field in a Conducting Atmosphere', ibid. **29 A**, No. 11.

—— 1943 *b*: 'On Sunspots and the Solar Cycle', ibid. No. 12.

—— 1945 *a*: 'Magneto-hydrodynamic Waves and Sunspots, I', *Mon. Not. R. Astr. Soc.*, **105**, 1.

—— 1945 *b*: 'Magneto-hydrodynamic Waves and Sunspots, II', ibid. p. 382.

—— 1947: 'Granulation, Magneto-hydrodynamic Waves, and the Heating of the Solar Corona', ibid. **107**, 211.

—— 1948: 'Magneto-hydrodynamic Waves and Sunspots, III', *Ark. f. mat., astr. o. fysik*, **34 A**, No. 23.

ALLEN, C. W., 1947: 'Interpretation of electron densities from corona brightness', *Mon. Not. R. Astr. Soc.*, **107**, 426.

APPLETON, E. V., 1945: 'Departure of Long-wave Solar Radiation from Blackbody Intensity', *Nature*, **156**, 534.

ÅSTRÖM, E., 1948: *Transactions of the Royal Institute of Technology, Stockholm*, **22**, 70.

D'AZAMBUJA, M. et Mme L., 1941: 'L'évolution et les mouvements d'ensemble des protubérances solaires'. *L'Astronomie*, octobre, 1941 (*Bull. de la Soc. Astr. de France*).

BABCOCK, H. W., 1947: 'Zeeman Effect in Stellar Spectra', Mt. Wilson Contr. No. 727, *Astrophys. J.*, **105**, 105; 'Remarks on Stellar Magnetism', *Publ. Astr. Soc. Proc.*, **59**, 112.

BIERMANN, L., 1946: 'Zur Deutung der chromosphärischen Turbulenz und des Exzesses der UV-Strahlung der Sonne', *Naturwiss.*, **33**, 118.

BJERKNES, V., 1926: 'Solar Hydrodynamics', *Astrophys. J.*, **64**, 93.

BLANCH, G., LOWAN, A. N., MARSHAK, R. E., and BETHE, H. A., 1941: 'The Internal Temperature-density Distribution of the Sun', ibid. **94**, 37.

BRUCE, C. E. R., 1946: 'Solar Prominences', *Observatory*, April 1946, p. 263.

CHAPMAN, S., and FERRARO, V. C. A., 1929: 'The Electrical State of Solar Streams of Corpuscles', *Mon. Not. R. Astr. Soc.*, **89**, 470.

COWLING, T. G., 1929: 'On the Radial Limitation of the Sun's Magnetic Field', ibid. **90**, 140.

Cowling, T. G., 1945 *a*: 'On the Sun's General Magnetic Field', ibid. **105,** 166.
—— 1945 *b*: 'The Electrical Conductivity of an Ionized Gas in a Magnetic Field, with Applications to the Solar Atmosphere and the Ionosphere', *Proc. Roy. Soc.* A, **183,** 453.
—— 1946 *a*: 'The Growth and Decay of the Sunspot Magnetic Field', *Mon. Not. R. Astr. Soc.* **106,** 218.
—— 1946 *b*: 'Alfvén's Theory of Sunspots', ibid. 446.
Edlén, B., 1942: 'Deutung der Emissionslinien im Spektrum der Sonnenkorona', *Zs. f. Astrophys.*, **22,** 30.
—— 1945: 'The Identification of the Coronal Lines', *Mon. Not. R. Astr. Soc.*, **105,** 323.
Elsasser, W. M., 1939: 'On the Origin of the Earth's Magnetic Field', *Phys. Rev.*, **55,** 489. See also 'A Statistical Analysis of the Earth's Internal Magnetic Field', ibid. **60,** 876, 1941; 'Induction Effects in Terrestrial Magnetism', ibid. **69,** 106, 1946; **72,** 821, 1947.
Ferraro, V. C. A., 1937: 'The Non-uniform Rotation of the Sun and its Magnetic Field', *Mon. Not. R. Astr. Soc.*, **97,** 458.
Fox, Ph., 1921: 'The Rotation Period of the Sun; Differential Motion of the Flocculi', *Publ. Yerkes Obs.*, **3,** 162.
Galvenius, I., and Wold, H., 1948: 'Statistical Tests of H. Alfvén's Theory of Sunspots', *Ark. f. mat., astr. o. fysik,* **34 A,** No. 24.
Giovanelli, R. G., 1948: 'Emission of Enhanced Microwave Solar Radiation', *Nature,* **161,** 133.
Grotrian, W., 1934: 'Über das Fraunhofersche Spektrum der Sonnenkorona', *Zs. f. Astrophys.*, **8,** 124.
Hale, G. E., 1908: 'On the Probable Existence of a Magnetic Field in Sunspots', Mt. Wilson Contr. No. 30, *Astrophys. J.*, **28,** 315; 1913: Mt. Wilson Contr. No. 71, ibid. **38,** 27.
—— Seares, F. H., van Maanen, A., Ellerman, F., 1918: 'The General Magnetic Field of the Sun', Mt. Wilson Contr. No. 148, ibid. **47,** 206.
Hey, J. S., 1946: 'Solar Radiation in the 4–6 Metre Radio Wave-length Band', *Nature,* **157,** 47.
Houtgast, J., 1947: 'Shock waves in the solar atmosphere', *Colloques Internationaux du Centre National de la Recherche scientifique, IX* (Lyon, Sept. 1947).
van de Hulst, H. C., 1947: 'Zodiacal Light in the Solar Corona', *Astrophys. J.*, **105,** 471.
Jansky, K. G., 1933: 'Electrical Disturbances, apparently of Extraterrestrial Origin', *Proc. Inst. Rad. Eng.*, **21,** 1387.
Kiepenheuer, K. O., 1946: 'Origin of Solar Radiation in the 1–6 Metre Radio Wave-length Band', *Nature,* **158,** 340.
Lundquist, S., 1948: 'On the Angular Momentum of the Sun', *Ark. f. mat., astr. o. fysik,* **35 A,** No. 27.
Martyn, D. F., 1946 *a*: 'Polarization of Solar Radio-frequency Emission', *Nature,* **158,** 308.
—— 1946 *b*: 'Temperature Radiation from the Quiet Sun in the Radio Spectrum', ibid. p. 632.
—— 1947: 'Origin of Radio Emissions from the Disturbed Sun', ibid. **159,** 26.

McCrea, W. H., 1929: 'The Mechanism of the Chromosphere', *Mon. Not. R. Astr. Soc.*, **89**, 718.

McCready, L. L., Pawsey, J. L., and Payne-Scott, R., 1947: 'Solar Radiation at Radio Frequencies and its Relation to Sunspots', *Proc. Roy. Soc.* A, **190**, 357.

Milne, E. A., 1924: 'The Equilibrium of the Calcium Chromosphere', *Mon. Not. R. Astr. Soc.*, **85**, 111; **86**, 578.

—— 1927: 'Selective Radiation Pressure and the Structure of a Stellar Atmosphere', ibid. **87**, 697.

Newton, H. W., 1934: 'The Sun's Rotation Period derived from Recurrent Sunspots', ibid. **95**, 60.

Öhman, Y., 1947: 'Results from Observations of the total Solar Eclipse of 1945, July 9', *Stockholms Obs. Ann.*, **15**, No. 2.

Payne-Scott, R., Yabsley, D. E., and Bolton, J. G., 1947: 'Relative Times of Arrival of Bursts of Solar Noise on Different Radio Frequencies', *Nature*, **160**, 256.

Pawsey, J. L., Payne-Scott, R., and McCready, L. L., 1946: 'Radiofrequency Energy from the Sun', ibid. **157**, 158.

Reber, G., 1944: 'Cosmic Static', *Astrophys. J.*, **100**, 279.

Rydbeck, O., 1948: 'On the Propagation of Waves in Inhomogeneous Media', *Trans. Chalmer's University of Technology*, **74** (Gothenburg).

Ryle, M., and Vonberg, D. D., 1946: 'Solar Radiation of 175 Mc./s.', *Nature*, **158**, 339.

Schwarzschild, M., 1946: 'On the Helium Content of the Sun', *Astrophys. J.*, **104**, 203.

—— 1948: 'On Noise arising from Solar Granulation', ibid. **107**, 1.

Shklowsky, J. S., 1947: 'Emission of Radio-waves by the Galaxy and the Sun', *Nature*, **159**, 752.

Southworth, G. C., 1945: 'Microwave Radiation from the Sun', *J. Franklin Inst.*, **239**, 285.

Strömgren, B., 1940: 'On the Chemical Composition of the Solar Atmosphere', *Festschrift für Elis Strömgren*, Publ. Københavns Observ. 127.

Thiessen, G., 1946: 'La Mesure du champ magnétique général du soleil', *Ann. d'astrophysique*, **9**, 101.

Vand, V., 1943: 'Temperature of Solar Corona', *Nature*, **151**, 728.

Waldmeier, M., 1939: 'Eigenbewegungen der Sonnenflecken', *Astr. Mitt.*, **14**, 453.

—— 1941: *Ergebnisse und Probleme der Sonnenforschung*, Leipzig.

—— 1945 *a*: 'Der physikalische Zustand der Sonnenkorona', *Mitt. Aarg. Naturforsch. Ges.*, **22**, 185.

—— 1945 *b*: 'Das Verhalten der Koronalinie 5694·42 Å', *Astr. Mitt., Zürich*, **147**.

—— 1946: 'Die Rotation der Sonnenkorona', ibid.

Walén, C., 1944: 'On the Theory of Sunspots', *Ark. f. mat., astr. o. fysik*, **30 A**, No. 15; ibid. **31 B**, No. 3.

—— 1946: 'On the Distribution of the Solar General Magnetic Field and Remarks concerning the Geomagnetism and the Solar Rotation', ibid. **33 A**, No. 18.

Wildt, R., 1939: 'Negative Ions of Hydrogen and the Theory of Opacity of Stellar Atmospheres', *Astrophys. J.*, **90**, 611, 1939.

—— 1947: 'An Interpretation of the Heights of Lines in the Solar Chromosphere', ibid. **105**, 36.

Woolley, R., 1947: 'The Solar Corona', *Austr. J. of Science*, **10**, No. 2.

VI

MAGNETIC STORMS AND AURORAE

6.1. Introduction

The intimate connexion between magnetic storms and aurorae makes it necessary to consider them as two manifestations of the same phenomenon. During a storm a disturbance field ('D-field') is superimposed on the earth's normal magnetic field. The first serious attempt to study this field was made by Birkeland. Our present knowledge of its structure is mainly due to the investigations by Chapman. The reader is referred to excellent surveys by Chapman (1936) and Chapman and Bartels (1940).

The disturbances have three maxima, two very strong ones in the auroral zones ('polar disturbances') and a smaller one at the equator ('equatorial disturbance').

The auroral zones are situated at geomagnetic latitudes of about $\pm 68°$. A magnetic storm seems always to be accompanied by a display of the aurora at least in some part of the auroral zones. At great storms the auroral zones are displaced towards the equator, and occasionally an aurora is observed even at very low latitudes, but this is a rare phenomenon. Near the poles the auroral frequency is lower than in the auroral zones.

Investigations of the disturbance field near the auroral zones have shown that it may be attributed to currents in the upper atmosphere flowing along these zones at a height of a hundred or a few hundred kilometres above the earth's surface. The current is directed westwards on the morning side and eastwards on the evening side of the earth. Hence the currents bring positive charge from the day side to the night side of the earth, but from magnetic data no certain conclusion has been reached as to how the current circuit is closed.

When we go from auroral latitudes towards the equator, the disturbances decrease until we reach the neighbourhood of the equator where a new but far smaller maximum is reached. This 'equatorial disturbance' manifests itself as a decrease of the normal field, pronounced on the evening side but rather weak on the morning side. As judged from the shape it may be due to a ring current encircling the earth in the westward direction in the equatorial plane at some distance from the earth.

A storm may begin suddenly ('sudden commencement') or as a more or

less continuous increase in magnetic activity. During the first few hours the disturbance field has a special character ('initial phase'), distinct from the following 'main phase'. The latter continues one or two days, after which it slowly fades out.

Magnetic storms are closely correlated with solar activity, which indicates that their ultimate cause is some disturbance in the sun. They often begin about one day after a solar flare, which leads to the conclusion that the disturbance is transmitted from the sun to the earth by an emission of corpuscles travelling with such a speed (average $= 2.10^8$ cm. sec.$^{-1}$) as to cover the distance in one day. Storms have a tendency to recur with a period of 27 days, which equals the solar synodic rotation. Frequently a recurrent storm is associated with an active group of sunspots, the storms occurring about one day after the spot group has passed the solar meridian. In many cases, however, no observable phenomenon in the sun can be associated with the occurrence of a storm.

The first serious attempt to interpret magnetic storms and aurorae is Birkeland–Störmer's theory. A recent review of it is given by Hewson (1937). For the development of this theory Birkeland's famous 'terrella' experiment has been of essential importance (Birkeland, 1908, 1913). In a big glass box, which was pumped out to the best vacuum attainable at that time, a magnetized steel sphere ('terrella') was placed. Between a cathode and the terrella (being an anode) a discharge was started. The discharge produced circular luminous rings around the magnetic poles of the terrella.

What really happened in the discharge no one knows. The experiment was carried out before the theory of gaseous discharges was developed. Birkeland and Störmer interpreted it in the following way. The cathode emits electrons ('cathode rays') which under the influence of the magnetic dipole field of the terrella move in complicated orbits later calculated by Störmer (see § 2.5). Störmer was able to show that electrons emitted from a source near the equatorial plane of the terrella should hit its surface in regions encircling the poles. The results were applied to nature: the sun was supposed to emit cathode rays which hit the earth, corresponding to the terrella, in the auroral zones around the poles.

This interpretation is probably not tenable. It is more likely that Birkeland's luminous rings were produced in the same way as a similar phenomenon in Malmfors's experiment, which we shall discuss in § 6.3.

In Birkeland's first papers the intimate connexion between the magnetic disturbances and the aurora was strongly emphasized, but during the development, including Störmer's mathematical theory and his

extensive auroral observations, only little attention was given to the magnetic side.

The main objections against Birkeland–Störmer's theory are the following. Electrons moving in Störmer orbits would hit the earth at the polar distance ($\sim 22°$) of the auroral zone if their energy were of the order of 10^8 e.v. The penetration power calculated from the minimum height of the aurora (80–100 km.) indicates, however, that the energy is only 10^4–10^5 e.v. Electrons with this energy would reach the earth only one or two degrees from the poles. Further, the theory does not take account of the solar magnetic field, which in fact had not been discovered in Birkeland's time. It is essential for the theory that this field should be zero, or in any case many orders of magnitude below the real value.

Finally, as first pointed out by Schuster, electrically charged particles of one sign would cause a prohibitively large space-charge if transmitted in sufficient number to cause terrestrial disturbances of the order of magnitude observed. This is a special case of the general rule that in cosmic physics the number of positive particles per unit volume must approximately equal the number of negative particles (see § 1.4). The conclusion is that the agent responsible for magnetic storms and aurorae must be a cloud which contains approximately the same amount of positive as of negative charge. Strong arguments in favour of this opinion have been given by Chapman and Ferraro (1931, 1932) and by Chapman and Bartels (1940). The cloud probably consists of ionized atoms.

The next attack on the problem was launched by Chapman and Ferraro (1931, 1932), who very carefully investigated the properties of an ion cloud emitted from the sun. (A recent review is given by Chapman and Bartels, 1940.) They find that space charge in the cloud is negligible except on the surfaces. If it travels with velocity v in a magnetic field H perpendicular to v it becomes polarized so that it is associated with a field E,

$$E = vH/c$$

(compare § 1.3). The electric lines of force are perpendicular to H as well as to v, and end on surface charges on both sides of the stream.

When the stream advances towards the earth, it is distorted by the terrestrial magnetic field. The interaction between the inertia of the stream and the forces induced by the magnetic field produces a system of currents. The magnetic disturbance which results from the theory has some similarity with the initial phase of a storm, but does not agree very well with the main phase disturbance. No immediate cause of the aurorae has been derived from the theory.

When discussing the polarization of the beam in a magnetic field, Chapman and Ferraro seem to consider especially the phenomena in the earth's magnetic field. Hence their electric polarization field is confined to the neighbourhood of the earth. Considering also the polarization in the solar magnetic field, a large-scale electric field is obtained with the same extension as the beam. (The breadth of the beam is of the order 5.10^{12} cm., whereas the region where the earth's magnetic field is of importance is $\sim 5.10^{10}$ cm.) Hence the entry of a beam into the neighbourhood of the earth means that the earth is placed in a general electric field. This is the starting-point of the 'electric field theory' (Alfvén, 1939, 1940), in which fundamental importance is attached to the electric field associated with the stream. (The inertia of the stream is more or less neglected.) Under the action of this field the particles of the stream drift in the magnetic field of the earth. The theory leads to a current system giving a magnetic field in rather good agreement with the observed disturbance field. The occurrence of the aurora is included in the theory. In fact, the aurora is supposed to take place where the current hits the atmosphere.

The arguments by which the current system was derived from the basic assumptions of the electric field theory were not compelling; the complexity of the problem makes an exact treatment impossible. Hence a scale-model experiment was made by Malmfors (1946). He put a terrella into an electric field and produced an ion stream drifting towards the terrella. A phenomenon was found which could be interpreted as an aurora, but this occurred also without the ion stream and it could be shown that the phenomenon was produced by an electric discharge. The conclusion was drawn that the properties of the stream are of little importance except for the production of the electric field. As soon as a sufficiently strong electric field is established in the environment of the earth a gaseous discharge occurs and the current system of the discharge is responsible for the magnetic disturbance. Aurora is produced in the regions where the discharge hits the earth.

Although a definite theory will probably include many properties of the stream as found by Chapman and Ferraro, the most promising approach at present seems to be to regard magnetic storms and aurorae as results of an electric field causing a gaseous discharge in the terrestrial magnetic field.

After this short general survey we intend to discuss the drift theory in § 6.2 and Malmfors's experiment in § 6.3. For a detailed account of the observational and theoretical results of the field the reader is referred to Chapman and Bartels (1940).

6.2. The electric field theory

The essential feature of the electric field theory of magnetic storms and aurorae is the assumption that the charged particles of the stream drift under the action of an *electric field* which is perpendicular to the solar magnetic field and to the motion of the stream. In the light of Malmfors's experiment the drift theory can probably not be upheld in its original form; still it is likely to contain many elements which should be included in a final theory. It is here presented essentially in its original form but with some modifications.

Let us consider the following simple case. The solar magnetic field H_0 is supposed to be homogeneous in the environment of the earth and parallel to the z-axis of an orthogonal reference system. The earth's dipole is situated at the origin of the system and is also parallel to the z-axis. An electric field E acts in the x-direction. In the solar magnetic field far away from the origin electrons move parallel to the xy-plane. The motion is treated according to § 2.3. Their energy is eV_0 and their orbital magnetic moment $\mu = eV_0/H_0$. Space charge is neglected. We also neglect the inertia term f^i in 2.3 (32). This means that we assume the circular velocity v to be large in comparison to the drift velocity $u_\perp$.

Under the influence of the electric field the electrons drift in the $-y$-direction. Those electrons which are in the neighbourhood of the $+y$-axis approach towards the earth. The electrons in the xy-plane, which is the earth's equatorial plane, move in the magnetic field H.

$$H = H_0 + \frac{a}{(x^2+y^2)^{\frac{3}{2}}}, \tag{1}$$

where a = the earth's dipole moment. The energy of an electron at (x, y) is given by

$$eV = eV_0 + eE(x_0' - x), \tag{2}$$

where x_0' gives its position far away from the earth. During the drift $\mu = eV/H$ will remain constant (see § 2.3). Hence the electrons move in small circles the centres of which drift along paths given by

$$\frac{eV_0}{H_0} = \mu = \frac{eV_0 + eE(x_0' - x)}{H_0 + a(x^2+y^2)^{-\frac{3}{2}}} \tag{3}$$

or

$$x_0 - x = L^4 r^{-3}, \tag{4}$$

with

$$r = \sqrt{(x^2+y^2)} \tag{5}$$

and

$$L = (\mu a)^{\frac{1}{4}}(eE)^{-\frac{1}{4}}. \tag{6}$$

The curves corresponding to different values of x_0 are shown in Fig. 6.1. It is evident that electrons coming from the sun pass on the morning

side if their x_0 value is larger than a certain critical value x_D^-, but on the evening side if $x_0 < x_D^-$. Consequently a stream of electrons becomes divided at a certain point x_d^-. The morning branch passes the earth at a rather large distance, but the evening branch encircles it and passes at a

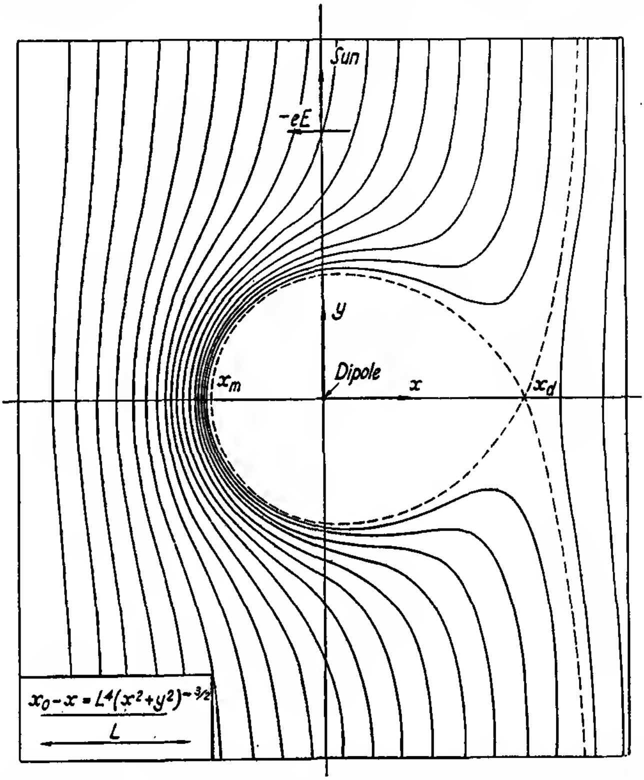

FIG. 6.1. Motion of the electrons in the equatorial plane of a magnetic dipole field on which is superimposed a homogeneous magnetic field, under the action of a homogeneous electric field.

minimum distance x_m^- from the dipole. A simple calculation gives from (4):

$$x_D^- = 4\sqrt[4]{3}\,L/3 = 1{\cdot}76L, \tag{7}$$

$$x_d^- = \sqrt[4]{3}\,L = 1{\cdot}32L, \tag{8}$$

$$x_m^- = -0{\cdot}74L. \tag{9}$$

Between the two branches there is a 'forbidden region'. The size of this region is determined by the value of the length unit L which depends on μ, E, and a in accordance with (6). The limit of the forbidden region is indicated by a dotted line in Fig. 6.1. Its equation is

$$\tfrac{4}{3}\sqrt[4]{3}\,L - x = L^4 r^{-3}. \tag{10}$$

The magnetic field produced by the drifting electrons is composed of the field of the dipoles μ (due to the electronic motion in small circles) and the field from currents (of negative charge) along the paths in Fig. 6.1. It can be shown that the latter effect predominates (see § 6.24).

The analysis of observational data of magnetic storms has shown that the disturbance field (D-field) can be divided into a polar disturbance and an equatorial disturbance (see, for example, Chapman and Bartels, 1940). The former can be attributed to currents in the upper atmosphere of the auroral zone: an afternoon current flowing eastward, and a morning current westward. The equatorial disturbance may also be due to currents in the upper atmosphere as postulated in Chapman's empirical current system, but most usually it is attributed to a ring current flowing westward in the equatorial plane and encircling the earth at some distance. As the equatorial disturbance is stronger in the afternoon than in the morning, the ring current is likely to be closer to the earth or stronger (or both) at the afternoon side.

From Fig. 6.1 it is evident that *the drifting electrons are equivalent to a ring current of about the type needed to explain the equatorial disturbance.* The current flows in the right direction; it is strongest and goes most closely to the earth on the afternoon side.

The ion cloud contains not only electrons but also positive ions. If the motion of the ions is treated in the same way as above, we find that they drift along curves which are the mirror images of the curves of Fig. 6.1, but whether enlarged or diminished depends upon the value of L for ions. As a and E are constants, the relative values of L for electrons and for ions depend only on μ which is proportional to the energy, i.e. temperature of the electronic and ionic gases before the stream enters the earth's field. If the electrons and ions of the stream are in temperature equilibrium, L is the same for both sorts of particles. On the other hand, if the stream is heated by the action of the electric field, we should expect plasma conditions as in a gaseous discharge. In this case the electronic temperature may be to the ionic temperature as 10, or a few powers of 10 to one (see § 3.13).

We assume that the ionic temperature is much lower than the electronic temperature. It must be admitted that this assumption is open to criticism. Hence the value of L for the ions is much smaller than for the electrons. This means that outside the forbidden region the ions move almost rectilinearly parallel to the y-axis.

As the velocity of the drift produced by the electric field is inversely proportional to H, and the drift produced by the inhomogeneity is

perpendicular to the magnetic gradient, it can easily be shown (Alfvén, 1940, p. 36) that if the inertia term f^i in 2.3 (32) is still neglected, the density of the particles is always proportional to H. Hence if initially the space charge in the stream is zero, as it must be according to § 6.1, no space charge occurs in those regions which are accessible to electrons as well as ions. This means that there is no space charge until we reach the forbidden region. On the day-side border of the forbidden region positive ions are accumulated. On the night-side border the electrons are uncompensated by the positive ions and a negative space charge is produced. *If we assume that these two regions of space charge neutralize each other by a discharge along the magnetic lines of force over the polar caps, we can account for the polar disturbance and the aurora.*

The gas pressure at the border of the forbidden region is certainly so low that the conductivity is much higher parallel to the magnetic field than perpendicular to it. Hence the current flows along the magnetic lines of force from the border-line until it reaches the upper atmosphere. We should expect the discharge to hit the upper atmosphere along a line which is the border-line projected along the magnetic lines of force upon the earth. *This line defines the auroral zone.*

6.21. *Auroral curve.* According to the theory, aurora should be observed where the current hits the upper atmosphere. This occurs along a line, called the *auroral curve*, which is the projection along the lines of force of the boundary of the forbidden region upon the earth. The boundary of the forbidden region is given, according to 6.2 (10), by

$$\tfrac{4}{3}\sqrt[4]{3}\,L - x = L^4 r^{-3}. \tag{11}$$

If a point on the auroral curve has the angular polar distance θ, and the longitude (or time-angle) λ, counted eastwards from the midnight point, we have [compare formula 1.2 (6)]

$$r \sin\lambda = x \tag{12}$$

$$r \sin^2\theta = R_0, \tag{13}$$

where R_0 is the radius of the earth (up to the upper atmosphere). Combining the equations (11), (12), and (13), we obtain the equation of the auroral curve:

$$\sin\lambda = \tfrac{4}{3}\sqrt[4]{3}\,(L/R_0)\sin^2\theta - (L/R_0)^4 \sin^8\theta. \tag{14}$$

Fig. 6.2 shows the curve (or rather that part of the curve of interest to us). At 6^h ($\lambda = 90°$) the polar distance θ has its smallest value,

$$\theta = \arcsin 3^{-\frac{1}{8}}\sqrt{(R_0/L)} = \arcsin 0{\cdot}872\,\sqrt{(R_0/L)}, \tag{15}$$

and at 18^h ($\lambda = 270°$) the maximum value

$$\theta = \arcsin 1{\cdot}164\sqrt{(R_0/L)} \tag{16}$$

is reached. The polar distance at a given hour depends upon the parameter $L = (\mu a/eE)^{\frac{1}{4}}$, and when this changes the curve becomes magnified or diminished. As the sine of the polar distance is proportional to the *eighth* root of E/μ, it is not very sensitive to changes in the field strength E or in the magnetic moment μ of the electrons. For example, if a point of the auroral zone has a polar distance of 20° for a certain value of E/μ and E/μ increases by a factor 5, the polar distance increases only $\sqrt[8]{5}$ ($= 1{\cdot}22$) times, corresponding to a displacement to 24·5°. This explains the relative constancy of the polar distances of the aurorae in spite of the fact that E and μ may fluctuate very much.

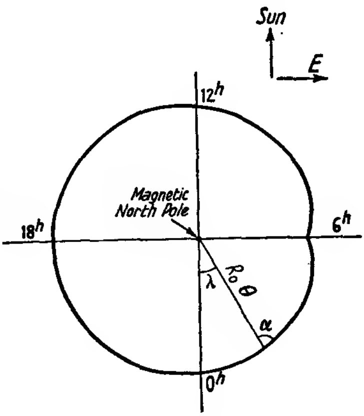

FIG. 6.2. Auroral curve (I-curve).

As it seems necessary to distinguish between different 'auroral curves', we call the curve defined by equation (14) *the ideal auroral curve* or *I-curve*. During the earth's daily rotation the geographical position of this curve is changed. If E/μ is constant and the earth's magnetic field is a dipole field, the curve rotates of course around the magnetic axis. The geographical path of an arbitrary point of the curve (given by its value of λ) during a day shall be called a *G-curve*. In the dipole case the G-curves are circles around the magnetic axis. The average of all the G-curves that correspond to the most frequent value of E/μ is the curve of maximum auroral frequency or *M-curve*.

If we take into account the difference between the earth's actual magnetic field and a dipole field, the conditions are more complicated. In this case the auroral curve at a certain instant, the *actual auroral curve* or *A-curve*, is still obtained by projecting the boundary of the forbidden region along the magnetic lines of force upon the surface of the earth (or rather upon the upper atmosphere), but the equation of the boundary as well as the equation of projection are now more complicated. But even if the actual field of the earth departs considerably from a dipole field at the earth's surface, at long distances from the earth it approximates more and more to a dipole field, because all terms of higher

order become negligible. Because of this we can assume that at the boundary of the forbidden region, which is usually situated at a distance of 5–10 times the radius of the earth, the magnetic field can be regarded as approximately a dipole field although at the earth's surface it departs considerably from a dipole field. Consequently, equation (1) still gives the boundary of the forbidden region.

Seen from the earth, an arbitrary point at the boundary of the forbidden region moves in a circle in the magnetic equatorial plane when the earth rotates. At the same time the corresponding point of the auroral zone follows a path which we have called a G-curve. Consequently, the G-curves are the projections of the circles in the equatorial plane along the magnetic lines of force upon the earth's surface. The M-curve (curve of maximum auroral frequency) can also be obtained in this way, because it has the character of a G-curve (corresponding to the mean distance of the boundary at the most frequent value of E/μ).

From auroral observations Fritz has computed the geographical situation of the M-curve. Vestine (1938, p. 261) has constructed the curve of maximum polar magnetic disturbances, which is very likely to be identified with the M-curve. On the other hand, the theoretical shape of the M-curve could be derived (as stated above) by computing the projection of a circle in the equatorial plane along the actual magnetic lines of force. If such a calculation is made, a comparison between the theoretical and observational curves would give a check of this theory.

To a first approximation the M-curve found by Vestine is a circle around the north pole of the eccentric dipole introduced by Bartels (1936) (see § 1.2). To the same degree of approximation the A-curve could be considered as an I-curve given by equation (14) if θ means the distance to Bartels's pole and λ the time angle with reference to it.

6.22. *Diurnal variation of the position of the aurorae.* If E/μ is constant during a day, an observer in the neighbourhood of the auroral zone must find a daily variation of the position of the aurorae. The polar distance of the auroral curve is a maximum at 18^h ($\lambda = 270°$) and a minimum at 6^h ($\lambda = 90°$). It is of interest to calculate the theoretical amplitude of this variation. Suppose that E/μ has such a value that an observer at a polar distance $\theta = 22°$ has the zone at the zenith at 0^h and 12^h ($\theta = 22°$ for $\lambda = 0°$ or $180°$). At 18^h the polar distance of the zone is 23·4° and at 6^h it amounts to 17·3° (values found from (4) by numerical calculation). Consequently its situation changes from 1·4° to the south to 4·7° to the north of our observer. Let us assume that there are aurorae

only just along the auroral curve (I-curve or A-curve), and that this curve is situated at a height of about 120 km. above the earth, which is the mean height of aurorae. Then a simple calculation shows that our observer will find the aurorae at 18^h at a height of about 37° above the southern horizon, at 0^h and 12^h at the zenith, and at 6^h about 11° above the northern horizon.

Of course the aurorae are usually spread considerably on both sides of the mathematical curve and, moreover, the polar distance of this curve changes always with E and μ. But still we should expect that because the mean position of the aurorae varies during the day, the ratio between the number of aurorae in the northern half of the sky and in the southern half will vary.

A daily variation of this kind has actually been observed. Results from Godthaab (Tromholt, 1882; Paulsen, 1893), Cap Thordsen (Carlheim-Gyllenskiöld, 1886) and Kingua Fjord (Internationale Polarforschung, 1882–3) are given in Figs. 6.4–6.7, whereas Fig. 6.3 shows the daily variation in the polar distance of the auroral curve according to theory. As pointed out above, we should expect the same type of variation in the relative number of aurorae observed to the north of the zenith at a station in the auroral zone. The sharp minimum of the polar distance comes at 5^h at Godthaab and at Kingua Fjord, and at 10^h at Cap Thordsen instead of at 6^h as according to theory, and the flat theoretical maximum at 18^h seems to come a little too late in all the observational curves.† Still the agreement must be considered as satisfactory.

Quantitative observations of the diurnal variations of the polar distance would be of great value.

6.23. *The direction of the auroral arcs.* From the theoretical point of view we must expect that the discharge over the polar regions of the earth should start from the boundary of the forbidden region. If the discharge is quiet, so that no secondary phenomena occur, the charged particles hit the upper atmosphere just along the auroral A-curve. Consequently this line is illuminated, and, as it is situated in the upper atmosphere, it is seen by an observer at the earth as an illuminated arc. This gives a reasonable explanation of the auroral arc, which in fact is one of the most common auroral forms.

If we accept this identification we must expect the direction of auroral arcs at a certain point and at a certain hour to be the same as

† The abscissae of the observational values give the local astronomical time, which must be corrected to 'auroral' or 'magnetic' time. A curve published by Vestine (1938) seems to indicate that the difference is less than one hour at Godthaab as well as at Cap Thordsen and, consequently of little importance for the comparison.

FIG. 6.3.

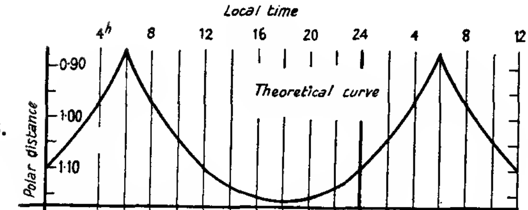

FIG. 6.4.

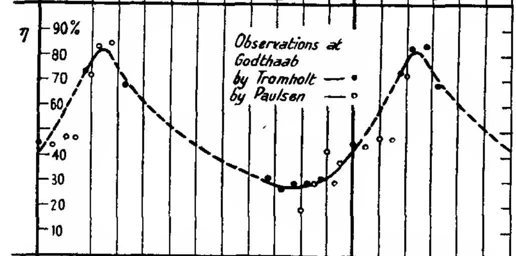

FIG. 6.5.

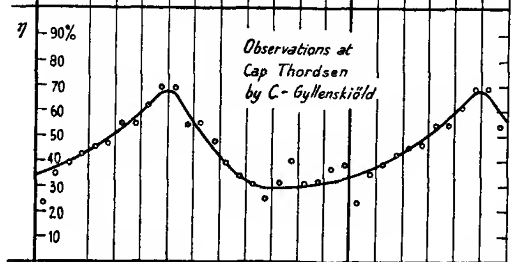

FIG. 6.6.

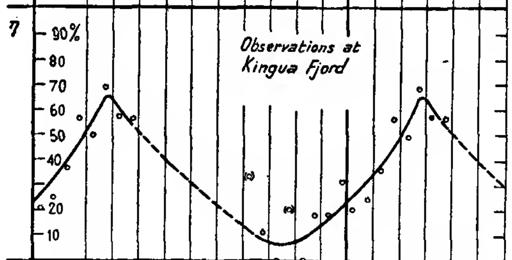

FIG. 6.7.

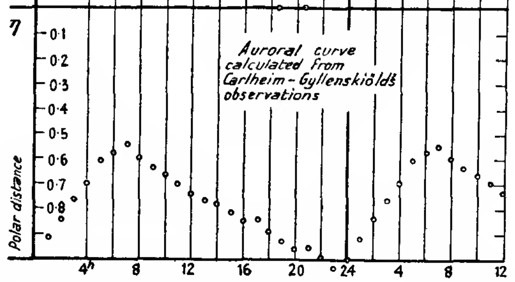

FIG. 6.3. Polar distance of the auroral curve according to theory.
FIGS. 6.4, 6.5, 6.6. Relative number of aurorae in the northern sky.
FIG. 6.7. Polar distance calculated from measurements of directions of arcs.

the direction of the auroral A-curve at that point and that hour. This gives us a possibility of calculating the direction of the auroral arcs at different times of the day. Neglecting the curvature of the earth (which is permitted when the polar distance θ is not too large) the azimuth α of

FIG. 6.8.

FIG. 6.9.

FIG. 6.10.

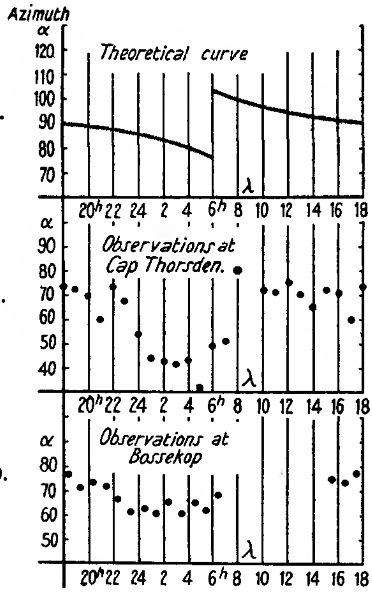

FIG. 6.8. Azimuth of auroral arcs according to theory.
FIGS. 6.9, 6.10. Azimuth of auroral arcs according to observations.

the direction of the auroral I-curve (compare Fig. 6.2) is given by

$$\cot\alpha = -\frac{d\theta}{\theta\, d\lambda}, \tag{17}$$

which can be calculated from 6.21 (14). We find

$$\cot\alpha = \frac{\cos\lambda}{8(L/R_0)^4\theta^8 - \frac{8}{3}\sqrt[4]{3}(L/R_0)\theta^2}, \tag{18}$$

which together with 6.21 (14) gives α as a function of λ. Fig. 6.8 shows how α varies with λ, i.e. how the azimuth of the auroral arcs varies with the time of day, according to the theory.

If the auroral A-curve is regarded as approximately an I-curve around the north pole of the eccentric dipole, the azimuth α given by (18) should be referred to the line from the station to this pole. Taking the eccentricity of the G-curves into consideration, we should rather refer it to a line perpendicular to the G-curve through the station.

The directions of the auroral arcs at different times of the day have been measured by Bravais at Bossekop and by Carlheim–Gyllenskiöld (1886) at Cap Thordsen. Their results are plotted in Figs. 6.9 and 6.10, which give the astronomical azimuth as a function of the local astronomical time (see the note on p. 185). The correction from astronomical azimuth to magnetic azimuth with reference to the north pole of Bartels's eccentric dipole is $-23°$ for Bossekop and $-37°$ for Cap Thordsen. If the magnetic azimuth is referred to the perpendiculars to the G-curves through the stations, the absolute values of the corrections are probably smaller (as judged from Vestine's M-curve).

Allowing for these uncertainties in the absolute value of the azimuth, there is no doubt a certain degree of accordance between the observational values and the theoretical curve. Although there is a considerable spread of the observational points—probably because the observational series are too short—the steady decrease of the azimuth from 6^h is clearly indicated at Cap Thordsen and also, to some extent, at Bossekop. The discontinuous rise at 6^h of the theoretical curve corresponds to a very steep rise from 32° at 5^h to 80° at 8^h in the values from Cap Thordsen. The amplitude of the variation has at Bossekop about the theoretical value; at Cap Thordsen it is somewhat larger.

Further observations are necessary, however, in order to make a definite check of the theory in this respect. In particular, observations at the discontinuity at 6^h are very important.

According to (17) the azimuth of the direction of the arc at a given hour is obtained by differentiating the auroral curve (I-curve or A-curve). Conversely we can construct this curve from observations of the directions of the arcs by integration. Assuming that the mean direction of the arcs at a certain hour (local time) is independent of the polar distance, we have in agreement with (17)

$$\frac{d\theta}{\theta} = -\cot\alpha\, d\lambda. \tag{19}$$

Suppose that the observational mean of α during the hour h to $h+1$ is α_h and the polar distance at the time h is θ_h. For $\frac{1}{2}\pi-\alpha \ll 1$, we have approximately

$$\Delta\theta = \theta_{h+1}-\theta_h = -\frac{\pi}{12}\theta_h \cot\alpha_h. \tag{20}$$

We do not know the exact correction, C, between the astronomical azimuth A_h, which is observed, and the 'auroral' azimuth α_h $(= A_h+C)$ which is to be used in our formula. If we put $C = \frac{1}{2}\pi-\bar{A}$, we obtain

$\cot \alpha_h = \tan(\bar{A}-A_h) \approx \bar{A}-A_h$. This gives

$$\theta_n = \theta_0 \prod_{h=0}^{n} \left[1-\frac{\pi}{12}(\bar{A}-A_h)\right]. \tag{21}$$

As we want to have $\theta_{24} = \theta_0$, this condition defines $\bar{A}$. $\bar{A}$ departs very little from the arithmetical mean of A_h.

In equation (21) we insert the azimuthal value for the auroral arcs as given by Carlheim-Gyllenskiöld. One missing value is interpolated. The curve obtained in this way is shown in Fig. 6.7. Comparing it with the theoretical curve of the polar distance (Fig. 6.3), we find that the sharp minimum of the polar distance comes at about the right place, but the maximum distance is attained too late. The amplitude of the variation is about twice the theoretical amplitude.

As a result of our comparison between the theoretical curve (Fig. 6.3) and the observational curves (Figs. 6.4, 6.5, 6.6, and 6.7) we may conclude that the observational data, although rather scarce, confirm the theoretical predictions about the form of the auroral curve.

6.24. *Current system of the magnetic disturbance.* We shall now analyse the current system of the theory in order to find the magnetic field of it. This field will later be compared with the observed disturbance field.

The motion of the charged particle in the stream is composed of the drift u^E due to the electric field and the drift u^m due to the inhomogeneity of the magnetic field. The former drift is the same for positive and negative particles, which means that it gives no resultant electric current in the ion cloud. On the other hand, the inhomogeneity of the magnetic field gives a drift which has opposite directions for positive and for negative particles, so that an electric current is produced in the ion cloud.

According to 2.3 (32) $u_\perp^m$ is perpendicular to f^m. As the current has the direction of $u_\perp^m$ and as f^m is parallel to the magnetic gradient which in the equatorial plane is antiparallel to the radius vector, the current flows in circles around the dipole.

When the ion stream invades the earth's magnetic field it is probably condensed to a rather thin disk in the magnetic equatorial plane. In this plane the invasion is stopped at a certain line which we have called the boundary of the forbidden region (see Fig. 6.1).

As the inhomogeneity of the earth's magnetic field gives an eastward drift to negative particles and a westward drift to positive particles, circular currents in the westward direction are produced in the ion cloud outside the boundary of the forbidden region (see Fig. 6.11). The ion density (number of ions per cm.^{2} of the equatorial plane) is proportional

to H or to r^{-3} (according to § 6.2) and the velocity of the drift is proportional to $H^{-1}\operatorname{grad} H \sim r^{-1}$. Hence the current density is

$$\frac{dI}{dr} = i_0 L^4 r^{-4}, \tag{22}$$

where i_0 is a constant. The constant L^4 is introduced in order to simplify subsequent equations.

As pointed out in § 6.2, we can expect the electrons in the equatorial

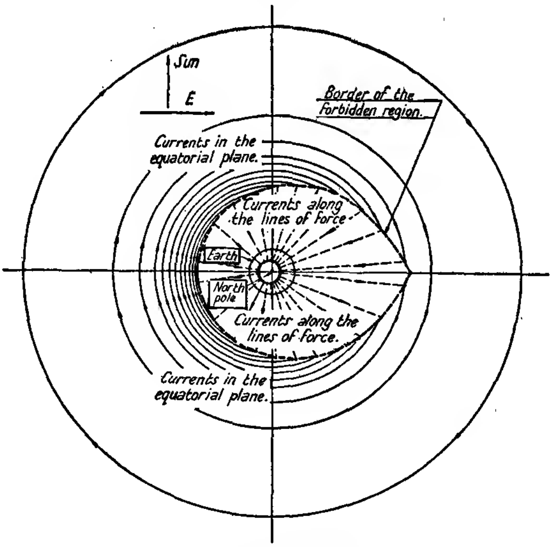

FIG. 6.11. Current system.

plane to give rise to the equatorial disturbance during a magnetic storm. The motion of an electron or ion is the resultant of a motion in a small circle and the drifts (discussed above) of this circle. Consequently the magnetic field is composed of the field h_1 from the small circular motion (which can also be considered as the diamagnetic action of the electron) and the field h_2 of the drift current. It can be shown that at the earth's surfaee $h_1 \approx -\frac{1}{3}h_2$. In the following h_1 will be neglected.

Thus the major part of the equatorial disturbance during a magnetic storm is due to the drift currents of the electrons and ions in the equatorial plane outside the forbidden region. As the earth's distance from the boundary of the forbidden region is a minimum at the evening side (18^h) and maximum at the morning side (6^h), the equatorial disturbance must attain a maximum at 18^h and a minimum at 6^h (see Fig. 6.11). The direction of the field at the earth's equator is southward and exactly

horizontal. All this is in good qualitative agreement with observational results.

A more detailed comparison will be given in § 6.25.

The currents in the equatorial plane flow along concentric circular arcs, which intersect the boundary of the forbidden region when their radii are smaller than the maximum distance of this line. Consequently they transport electric charge to the boundary of the forbidden region. It is easily seen that this polarization of the ion cloud supplies positive charge to the day-side of the boundary (6^h–18^h) and negative charge to the night-side of it (18^h–6^h). The space charge increases continuously until it becomes discharged in some way. As the gas density is very low, the parallel conductivity $\sigma_{\parallel}$ is very much greater than the cross-conductivity $\sigma_{\perp}$ (see § 3.21). Hence the current will leave the equatorial plane and flow along the magnetic lines of force until it reaches the upper atmosphere.

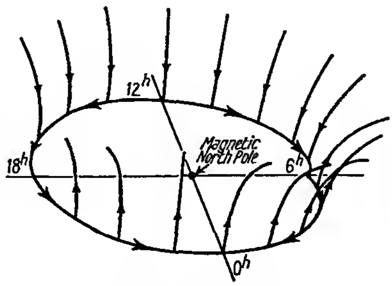

FIG. 6.12. Current system of the auroral zone.

The magnetic lines of force going through the boundary of the forbidden region intersect the surface of the earth (or rather the upper atmosphere) in points of the curve called the *auroral curve* given by 6.21 (14). Fig. 6.2 shows its form. If we assume this curve to be electrically conducting, the discharge currents along the lines of force continue their way along this curve. Consequently the positive charge at the day-side of the forbidden region is able to neutralize the negative charge at the night-side through currents which, first, follow the magnetic lines of force from the equatorial plane to the upper atmosphere of the north and south auroral curves, then flow along these curves from the day-side to the night-side of them, from where they go back to the equatorial plane along the magnetic lines of force (see Fig. 6.12). This current system, which is responsible for the *polar disturbance*, completes the circuit of the currents (discussed above) which produce the equatorial disturbance.

The assumption of high electrical conductivity along the auroral curve is very natural. As a consequence of the discharges along the magnetic lines of force, this curve is hit by charged particles, which ionize the upper atmosphere greatly. This is confirmed by ionosphere measurements. According to Harang (1937) the ionization in the

E-region (height $\approx$ 100 km.) at Tromsö, close to the auroral zone, increases even during small magnetic storms by a factor of 10 or more.

We can therefore expect the conductivity along the auroral curve to be much larger than in the environment, which has the consequence that the current prefers the auroral curve. The effect in § 5.62 is less pronounced because the dimensions are smaller and the conductivity lower, but should probably be considered.

The current density of the polar disturbance system can be computed in the following way. The equatorial current to an arc of the boundary of the forbidden region between the points (r, x) and $(r+dr, x+dx)$ amounts to

$$dI = i_0 L^4 r^{-4}\, dr \tag{23}$$

according to (22). But after differentiating equation 6.2 (10) we have $dx = 3L^4 r^{-4}\, dr$, which gives

$$dI = \tfrac{1}{3} i_0\, dx. \tag{24}$$

Consequently, the current leaving the equatorial plane along the lines of force from a given arc of the boundary of the forbidden region is proportional to the projection of this arc on the x-axis.

The current reaching the auroral zone between the longitudes λ and $\lambda+d\lambda$ is dI, given by

$$dI = \tfrac{1}{3} i_0 \frac{dx}{d\lambda}\, d\lambda, \tag{25}$$

In order to calculate $dx/d\lambda$ we put the relation $r = x/\sin\lambda$ into 6.2 (10). This gives

$$x^3(\tfrac{4}{3}\sqrt[4]{3}\, L - x) = L^4 \sin^3\lambda. \tag{26}$$

Differentiation of this equation gives $dx/d\lambda$.

The current along the magnetic lines of force goes from the equatorial plane to the northern as well as to the southern auroral zones. Although different conditions in the two hemispheres (e.g. summer with greater conductivity in one of them) might disturb the symmetry sometimes, it is reasonable to assume that, as a mean, half of the current goes either way.

Part of the current coming along the magnetic lines of force to the auroral curve is discharged over the evening branch of the curve, and part of it over the morning branch. It is difficult to predict on purely theoretical grounds how much goes either way. If the discharge along the auroral curve always chooses the shortest way, as is reasonable to assume, the current becomes divided at about the 0^h and 12^h points. This means that the positive charge coming in between the 12^h and 18^h points flows over the evening branch and neutralizes the negative charge reaching the auroral curve between the 18^h and 24^h points. In the same way the

positive charge between 6^h and 12^h flows over the morning branch to the 0^h–6^h part of the curve.

On the assumption made above we can calculate the current $I(\lambda)$ at a point of the auroral curve with the longitude λ (counted eastward from the midnight direction). The difference $(dI/d\lambda)\,d\lambda$ between the current at the longitudes λ and $\lambda+d\lambda$ equals the inflow of current along the magnetic lines of force, which is given by (25). If $I(\lambda)$ represents the current

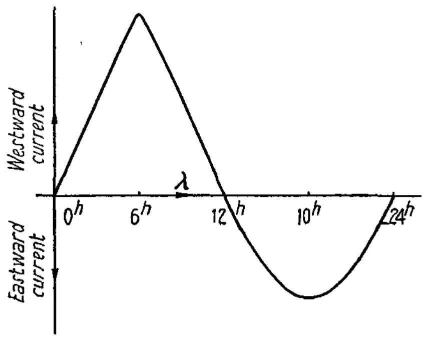

FIG. 6.13. Current along the auroral zone.

in only one of the two auroral curves (the northern or the southern curve) we must divide the inflow by 2. After integration we have

$$I(\lambda) = \tfrac{1}{2}\int\limits_0^\lambda \frac{dI}{d\lambda}\,d\lambda = \tfrac{1}{6}i_0\int\limits_0^\lambda \frac{dx}{d\lambda}\,d\lambda,$$

or

$$I(\lambda) = \tfrac{1}{6}i_0\,x, \tag{27}$$

where x is given as a function of λ by (26). Fig. 6.13 shows how I depends upon λ. The ratio between the maximum westward current (at 6^h) and the maximum eastward current (at 18^h) equals the ratio x_d/x_m between the maximum and minimum distances of the forbidden region from the origin. According to 6.2 (8) and 6.2 (9) this ratio is

$$1{\cdot}32 : 0{\cdot}74 = 1{\cdot}78.$$

6.25. *Magnetic field of the current system.* The theoretical current system is defined by equations 6.24 (22), 6.24 (24), and 6.24 (27). The next task is to find the magnetic field which it produces. This can be done by ordinary mathematical methods, but as this requires very much numerical integration it was preferred to make a scale model of the system and measure the field with a search coil. Two models (scale $1:10^8$) were constructed; in the 'first model' the height of the currents in the auroral zone over the earth corresponded to about 500 km., in the

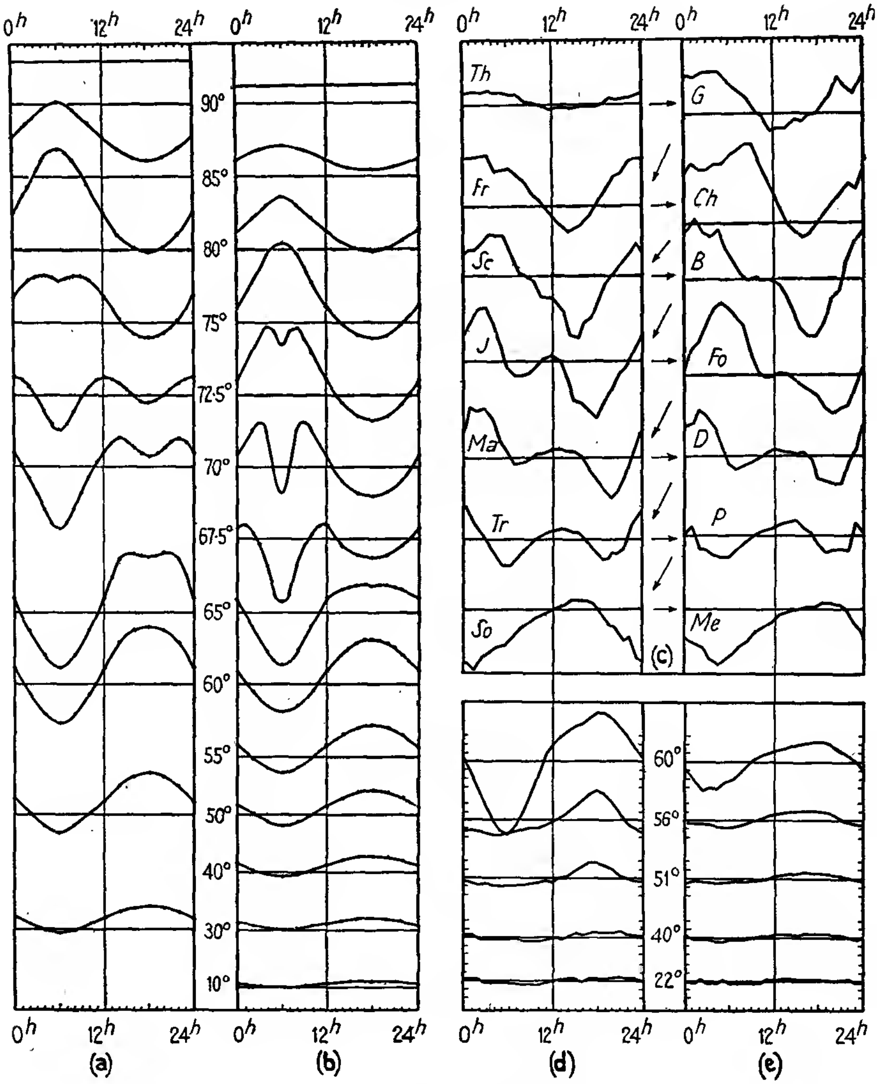

FIG. 6.14. Vertical component of disturbance. (*a*) and (*b*) theoretical curves, (*c*), (*d*), and (*e*) observations.

'second model' to 300 km. Even this height is perhaps larger than what should be expected from observations. The minimum and maximum polar distances of the auroral curve were 18° 45′ and 25° 20′ in approximate agreement with the polar distance of the zone of maximum auroral frequency (22–23°).

The vertical, north–south, and east–west components of the magnetic

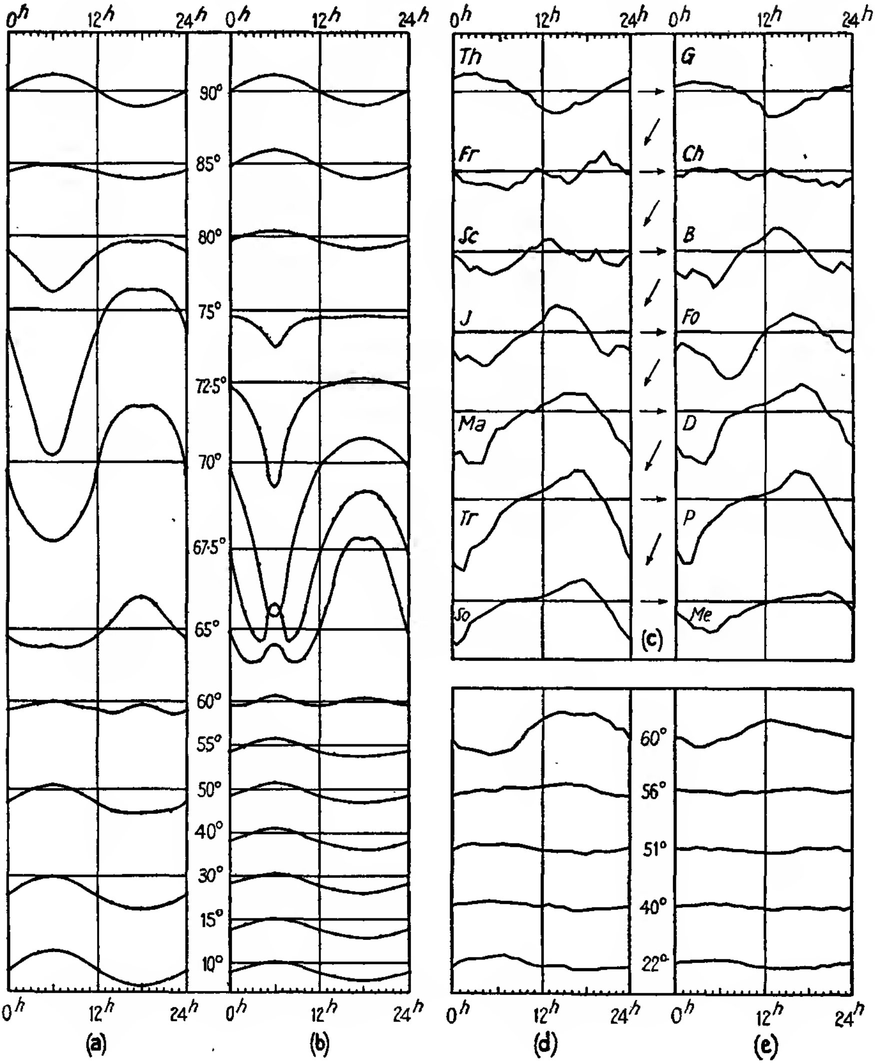

FIG. 6.15. South–north component of disturbance.

field of the current system were measured by a search coil at different points of a sphere corresponding to the surface of the earth. The search coil was set at a given latitude and readings were taken at different longitudes. The curves obtained in this way for different latitudes give the *diurnal variation of the magnetic storm field* according to this theory.

The result of the measurements is shown in Figs. 6.14, 6.15, and 6.16*a*

and b, which give the three components of the *theoretical disturbance field* as found with the first (a) and the second (b) models.

The curves obtained from the models are to be compared with the *observational curves*, compiled from averages over a number of storms

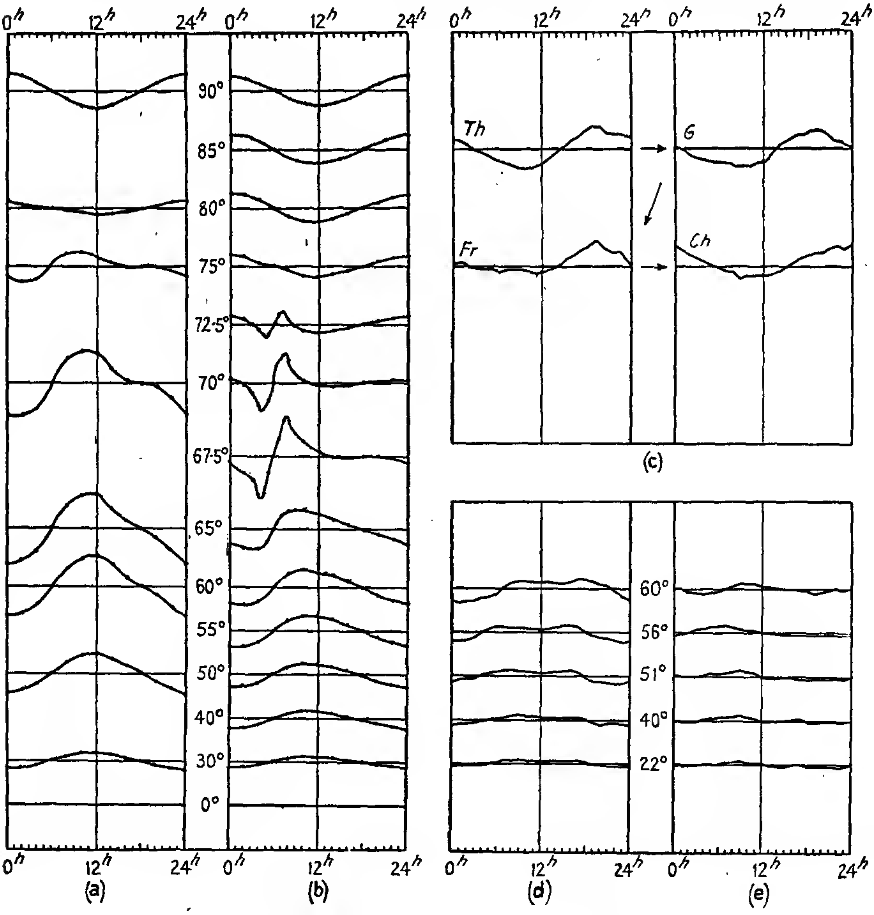

FIG. 6.16. East–west component of disturbance.

or disturbed days. Figs. d and e of Figs. 6.14, 6.15, and 6.16 show S_D, the diurnal variation of the disturbance at low and temperate latitudes, as found by Chapman (1919 and 1927).

Figs. c of Figs. 6.14, 6.15, and 6.16 are obtained from data given by Vestine and Chapman (1938) and represent the disturbance field at polar latitudes. The curves are obtained by adding the diurnal variation S_D to the daily mean D_m of the disturbance, so that they represent the

total disturbance field. The vertical component is taken from Vestine and Chapman's curves, the north–south and east–west component from their vector diagrams. As the east–west component in the auroral zone is small in comparison with the north–south component, its values are not very reliable, and comparison with the theoretical values would have no significance.

A comparison between the theoretical curves obtained from the experiments upon the model and the observational curves shows a good general agreement, the main discrepancy being that the observational curves are sometimes displaced a few hours in relation to the theoretical curves.

The *vertical force* has the following types of variation according to Fig. 6.14:

1. At the pole it has a constant positive value.

2. For stations situated between the pole and the minimum polar distance of the auroral curve the vertical force varies approximately as a sine curve with a period of 24^h and a maximum at 6^h. The amplitude increases with increasing polar distance.

3. Stations situated to the south of the minimum polar distance of the auroral curve, but to the north of the maximum polar distance of it, have approximately a *double sine curve* (period 12^h) with maxima at 0^h and 12^h. This is due to the fact that these stations are to the north of the auroral curve at 18^h but to the south of it at 6^h. Consequently their curves show the minimum at 18^h which is characteristic for stations to the north of the auroral zone, but at the same time the minimum at 6^h which is found at stations to the south of the auroral zone.

4. Stations definitely to the south of the auroral zone have approximately a sine curve with a 24^h period and a maximum at 18^h. The amplitude decreases with decreasing latitude.

5. At the equator the vertical disturbing force is zero because of the symmetry of the current system.

The accordance between the theoretical curves (from the model) and the observational curves is especially striking within the auroral zone, where the simple sine wave changes into a double wave. This type of variation seems to have attracted little interest, although it is very distinct in Vestine and Chapman's observational curves. The current system constructed by Chapman to reproduce the observations does not give this type of variation. On the other hand, the double wave follows immediately from the electric field theory and is due to the eccentricity of the auroral curve. Thus the fact that double-waves are

really observed within the auroral zone adds further weight to the arguments given in §§ 6.21, 6.22, and 6.23 for the view that the auroral curve (I-curve) has the eccentric form predicted by theory.

The *south–north component* of the disturbing force exhibits the following types of variation (see Fig. 6.15).

1. In the neighbourhood of the pole it varies as a sine curve with the maximum at 6^h.

2. About half-way between the pole and the auroral zone the phase changes, so that the maximum occurs at 18^h in the vicinity of the auroral zone. The amplitude is largest at the zone of maximum auroral frequency.

3. Somewhat to the south of the auroral zone the phase changes again and we obtain the equatorial type of variation. The maximum occurs at 6^h. The amplitude increases with decreasing latitude.

Also the south–north component curves show a satisfactory agreement between theory and observations.

The diurnal variations of the *east–west component* are of the following types (see Fig. 6.16).

1. To the north of the auroral zone it varies as a sine curve with its maximum at 0^h.

2. Within the auroral zone it is difficult to compare the theoretical and observational curves for two reasons: the theoretical curves depend rather much on the height of the auroral zone currents, as shown by the difference between the curves a and b; the observational values are uncertain because of the asymmetry of the earth's magnetic field. As this differs from a dipole field, the directions in which we have to count the south–north and the east–west components do not always coincide with the directions of the magnetic meridian and the perpendicular to it. If we still take the magnetic disturbance force perpendicular to the magnetic meridian as the east–west component, we risk getting a fraction of the south–north component superimposed on the real east–west component. The south–north component being much larger than the east–west component, the east–west curves may become completely distorted.

If we assume that the maximum disturbance vector (in the horizontal plane) marks the direction of the south–north component and take the component perpendicular to it as east–west component, we obtain curves which are similar to the curves b (second model). The deflexions some hours before and after 6^h are not so large, however, as in the theoretical curves.

As a comparison of this kind must be somewhat arbitrary, the observational curves from the auroral zone are omitted.

3. To the south of the auroral zone the east–west component varies approximately as a sine curve. The maximum occurs at 12^h, and the amplitude decreases with decreasing latitude. The observational curves are a little irregular, a double maximum occurring in several cases.

4. At the equator the east–west component is zero because of the symmetry.

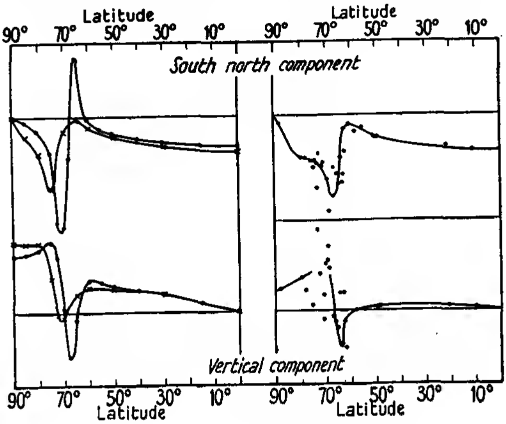

FIG. 6.17. Daily mean of disturbance.

Upon the whole, the agreement between theory and observations is rather satisfactory.

It is also possible to compare the experimental results from the model with observations in another way.

We calculate the average value D_m of each of the magnetic elements at a certain latitude as measured upon the models (D_m is the mean of the values at $0^h, 2^h, \ldots, 22^h$). The values of the vertical and south–north components at different latitudes are shown in Fig. 6.17 *a*. The D_m values of the east–west component are always zero because of the symmetry.

These theoretical curves are to be compared with the observational curves of the same quantity (see Fig. 6.17 *b*). The general agreement is satisfactory. The very sharp maxima and minima of the curves for the second model are not found in the observational curves, but this was not to be expected, because these are averages from observations of many storms having different values of the polar distance of the auroral zone. The observational curves must therefore be more smooth.

6.26. *Discussion of the electric field theory.* The attractive features of the theory are essentially the following.

1. The introduction of the electric field gives the problem that symmetry with respect to the 6^h–18^h line which is a characteristic feature of the disturbance field.

2. The aurora and the magnetic disturbances can be regarded in a natural way as two aspects of the same phenomenon.

3. A current system is obtained which gives a D-field in good agreement with the observed field.

4. An eccentric auroral line is obtained so that the diurnal variation of the position of the aurora is explained.

On the other hand, the following objections can be made:

1. The theoretical deductions are not stringent. Because of the complexity of the phenomena an absolutely stringent analysis is hardly possible.

2. Cowling (1942) has objected that if the particles of the stream move with μ ($= eV/H$) constant, they must have enormous energy in order to reach the upper atmosphere where H is large. Of course the condition μ = const. holds only as long as interaction between the particles does not change the velocity vector. It is quite reasonable to assume that in connexion with the discharge along the magnetic lines of force secondary effects occur (e.g. some type of plasma oscillations), so that the direction of the velocity vector changes.

3. As Cowling (1942) has pointed out, the electrons would never reach the vicinity of the border-line of the night side. Instead they should be discharged as soon as they leave the region which the ions can reach along straight paths. This objection may in part be correct. Of course it would be possible for the electrons to go all the way along the night side of the forbidden region if only enough positive ions were supplied by the discharge along the magnetic lines of force. But it is more reasonable to suppose that at least the main part of them are discharged before they have travelled the whole way. This calls for a modification of the theory, but this modification is very welcome from the observational point of view, because the night maximum of the aurora does come much earlier than required by the unmodified strictly symmetrical theory. Also, the magnetic disturbances are displaced in the direction required by this objection.

4. The inertia term cannot be neglected for the ions, because before they have reached the earth their drift velocity is likely to be larger than their circular velocity. Hence space charge is produced even outside the border line.

5. The most important objection is certainly that Malmfors's scale-

model experiment indicates that 'aurora' is produced independently of the ion stream by a discharge produced by the electric field. After a review of his experiment we shall return to the discussion of the theory.

6.3. Malmfors's scale-model experiment

The phenomena occurring during a magnetic storm are too complicated to be analysed in a purely theoretical way. Under such conditions a scale-model experiment may give important information. Of course one must be careful in applying the model results to nature.

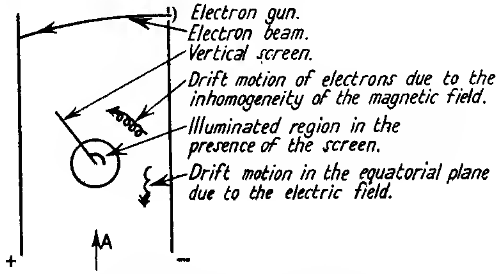

FIG. 6.18. Malmfors's terrella experiment.

Malmfors (1946) tried to reproduce the conditions postulated by the electric field theory: A terrella, i.e. a homogeneously magnetized steel sphere, representing the earth, was placed in an electric field between two parallel plates. This field should represent the field produced by the polarization of the ionized stream. The whole arrangement was placed in a vessel containing air at a pressure of 10^{-3} or 10^{-4} mm. The stream itself was produced by ionizing the gas between the plates in that direction from the terrella which, with regard to the sense of the electric and magnetic fields, represented the direction towards the sun. The ionization was produced by an electron beam which was curved (see Fig. 6.18) because of the magnetic field. The slow secondary electrons produced by this beam drifted in the combined electric and magnetic fields towards the terrella. The magnetic field was not strong enough to ensure that the ions also drifted in the same way, so in this respect the model does not resemble what, according to theory, occurs in nature. The similarity transformation in § 3.12 shows that it is beyond our experimental facilities to produce magnetic fields strong enough to correspond to these conditions. As the terrella is about 10^8 times smaller than the earth, the pressure (10^{-3}–10^{-4} mm.) corresponds to 10^{-11}–10^{-12} mm. or 10^6–10^5 particles/cm.^{3} This may be more than the normal density at the distance

of the 'border-line', but is much less than the density in the upper atmosphere at the normal height of the aurora. The applied voltage was some thousand volts, somewhat lower than the voltage (10^4–10^5 volts) expected over the forbidden region. Hence the electric field is not very much lower than that required by the similarity transformation.

Under the above condition eccentric luminous rings around the poles of the terrella were observed (see Fig. 6.19). Their maximum polar distance was in the direction corresponding to 18^h. Hence the rings

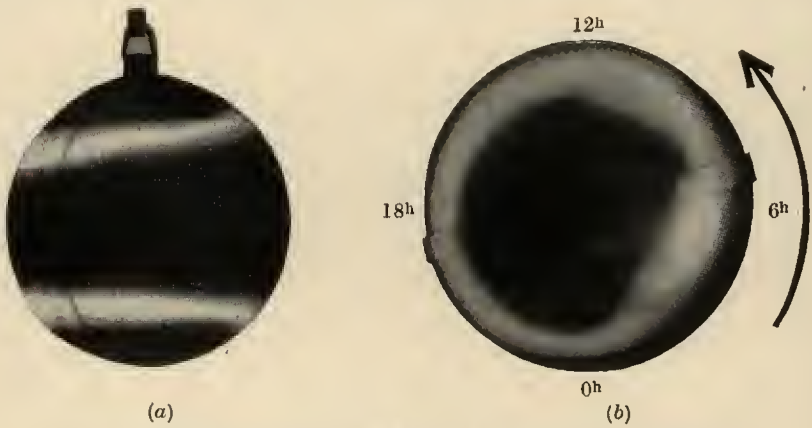

Fig. 6.19. Luminous rings on the terrella, seen from the 'night side' (*a*), and from the north pole (*b*).

could be identified with the aurora predicted by the theory. The current system has not been measured.

So far the experiment seemed to support the drift theory. Unfortunately for the theory, however, the same phenomena occurred even when the electron beam was put out of action so that no ionized stream was produced. Without an ionizer the gas pressure must be kept within narrower limits than with it.

If a 'vertical' screen is placed along a meridian of the terrella those parts of the luminous rings vanish which are situated between the screen and the 6^h point, reckoned anti-clockwise if seen from the north pole, whereas the parts between the 6^h point and the screen are unaffected (see Fig. 6.18). This indicates that the rings are due to electrons generated in a region in the 6^h direction and drifting anti-clockwise round the earth.

According to Malmfors the explanation of the phenomenon is likely to be the following. The pressure in the apparatus is so low that an electron moving rectilinearly from one electrode to the other does not ionize enough to produce a gaseous discharge. Only if its path is lengthened,

e.g. through a periodic motion, does its chance of producing enough secondaries become high enough. This is a well-known phenomenon: in a 'Penning manometer' (Penning, 1937) a glow discharge at much lower pressure than is normal is produced by the help of a magnetic field which compels the electrons to oscillate along the lines of force. In the terrella experiment electrons in a region ('generating region') in the 6^h direction can move in periodic orbits because the electric drift and the magnetic drift (due to the inhomogeneity of the magnetic field) compensate each other. In the electric field theory this occurs for the equatorial plane at x_d of Fig. 6.1. In the experiment the motion is probably complicated by oscillations along the magnetic lines of force. Anyhow in the generating region periodic orbits are possible, so that the conditions for a discharge at very low pressure are satisfied. Some of the electrons produced by the discharge leave the generating region and drift in the combined electric and magnetic fields around the earth. If the electronic motion were confined to the equatorial plane, they would follow the 'border-line' in Fig. 6.1. The oscillations along the lines of force, which are essential here, probably change the shape of the 'border-line' somewhat. As judged from the shape of the rings, its eccentricity seems to be greater than in the drift theory.

Malmfors's experiment resembles Birkeland's terrella experiment in certain respects (compare § 6.1). The luminous rings obtained on the terrella by Birkeland have been interpreted as due to cathode rays moving in Störmer orbits. This is probably not right. The photographs resemble those of Malmfors and the rings may have been produced by a gaseous discharge in an electric field near the terrella. If in Malmfors's experiment the field near the terrella is made approximately radial by connecting the two parallel plates together to a negative voltage and the terrella to a positive voltage, the rings become circular and symmetric around the poles, as in Birkeland's experiment.

Although we ought to be careful in applying a result from a scale-model experiment to nature, there seems to be no essential objection to the assumption that Malmfors's mechanism is active in the production of magnetic storms and aurorae. His theoretical treatment, although rudimentary, indicates that if enlarged to geophysical dimensions the mechanism would produce aurorae at about the right polar distance. A weak point is that the magnetic field of the model is too feeble to make ions spiral, but as ions usually move much more slowly than electrons, the ionic motion may not be essential. In most gaseous discharges the ionic motion in the plasma is not very important.

To sum up, no real conflict exists between Malmfors's experiment and the electric field theory and a modification of the theory guided by the experiment should be possible.

6.4. The aurora as an electric discharge

According to Birkeland–Störmer's theory the aurora is due to particles which hit the upper atmosphere and penetrate to a level given by the energy of the particles and the stopping power of the upper atmosphere. According to Störmer and Vegard electrons must have a primary energy of 10^4–10^5 volts in order to reach the lowest height (80–100 km.) at which the aurora is observed. If the primary cause of the aurora is an electric discharge, as indicated by the electric field theory and by Malmfors's experiment, this discharge may produce swift electrons outside the earth's atmosphere. When these electrons hit the atmosphere the same atmospheric phenomena as in Birkeland–Störmer's theory are produced. The voltage difference between the points x_m and x_d in Fig. 6.1 is, according to the electric field theory, of the order 10^4–10^5 volts, so the electrons incident upon the upper atmosphere would have the right energy.

There are, however, several arguments in favour of the view that the atmospheric phenomena should rather be treated as part of the general discharge. If so, a considerable part of the available voltage should be concentrated in the upper atmosphere where it produces a discharge in the atmosphere itself. This discharge carries current between space charges far away from the earth and some conducting layer (e.g. the E-layer) of the upper atmosphere. Plasma conditions, including the presence of an electric field, may prevail the whole way from the top of the atmosphere down to the lowest limit of the aurora. Hence no electrons with higher energy than some 10 volts need to be present. In this connexion it is of interest to note that, from comparison of laboratory spectra with the auroral spectrum, Bernard (1940) has drawn the conclusion that the aurora should be excited mainly by electrons of about 30 volts. His results probably do not exclude that there are primary electrons, with high energies, because even in this case the main excitation would be produced by slow secondary electrons. The present situation seems to be that spectroscopic data give no evidence for the existence of high-energy electrons nor for the absence of them.

A common auroral form is the auroral ray. A ray may be seen in the same place in the sky for a considerable time. This means that it is at rest in relation to an observer on the earth. If the aurora is considered

to be due to incident particles, an auroral ray must be produced by a thin bundle of electrons which are emitted from the sun (according to Birkeland–Störmer) or produced by a discharge near the earth. In both cases it is very difficult to understand why the point where the bundle hits the atmosphere should take part in the earth's rotation so as to make the ray immobile in relation to an observer on the earth. On the other hand, if the aurora is assumed to be an atmospheric discharge, the current can be expected to go the way where the resistance is least, i.e. the ionization a maximum. Hence if once a ray has been formed, the discharge has a tendency to continue along the same track. (Compare a flash of lightning, where several consecutive discharges follow the same path.) In this way the ray may remain immobile in relation to the atmosphere and to an observer on the earth's surface.

According to this 'atmospheric discharge' theory the dissolution of a continuous auroral arc† into a multitude of rays should be a phenomenon related to the constriction of a discharge.

As has been pointed out in § 3.4, constriction of a discharge may be due to thermal effects or to electromagnetic attraction. In the former case which occurs, for example, in an arc† at atmospheric pressure, the narrow channel to which the discharge is confined is heated by the current so much that the conductivity becomes larger than in the environment. In the latter case the electromagnetic attraction between parallel currents effects the constriction, and no significant rise in temperature is necessary. (In solar prominences the temperature in the discharge channel is lower than in its surroundings; see § 5.6.)

If auroral rays are interpreted as constricted discharges, the question whether the constriction is of thermal or electromagnetic origin could be solved if reliable temperature measurements were available. If a considerable difference in gas temperature between an auroral ray and the environment were found, this would speak in favour of the thermal constriction hypothesis, whereas if no temperature difference existed, the electrodynamic mechanism—or eventually some unknown mechanism—should be active. Measurements on band spectra indicate that the temperature at auroral heights is of the order of 200° K. (Vegard, 1939) and that in an auroral ray it does not increase with the height (Vegard and Kvifte, 1945). These results probably exclude the existence of thermal constriction.

Besides the constriction there is also another secondary phenomenon which may be of importance in producing the fascinating variety of

† Unfortunately both an auroral type and a discharge form is called an 'arc'.

auroral forms. Suppose that the curve of Fig. 6.20 *a* is a part of the auroral curve. If the discharge current is considerable and somewhat irregularly distributed, the potentials of different points along the curve may fluctuate. This holds not only for the auroral curve but also for the boundary line of the forbidden region and for the magnetic lines of force connecting them, i.e. along the whole path of the discharge. Let us consider the effect of an increase of the potential at a certain point *A* of the curve. From this point we have a radial electric field *f* in all directions. But in a magnetic field perpendicular to it, an electric field gives rise to a motion which is perpendicular to the magnetic field as well as to the electric field. Thus, if in Fig. 6.20 *a* the magnetic lines of force go up through the paper, the radial electric field around *A* will produce a vortical motion. Consideration shows that it goes in the clockwise direction if *A* has a positive charge and counter-clockwise if the surplus charge is negative. Consequently, a small space charge will cause a wave in the auroral arc like Fig. 6.20 *b*. If the space charge increases, this wave develops into forms like Fig. 6.20 *c* and 6.20 *d*. This corresponds to the drapery-shaped arcs and draperies, well-known auroral forms.

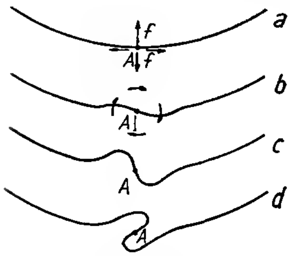

FIG. 6.20. Formation of draperies.

It seems likely that at least the greater part of the auroral forms could be considered as due to the effect of one or both of these secondary phenomena upon the auroral discharge.

REFERENCES

General references:

CHAPMAN, S., and BARTELS, J. 1940: *Geomagnetism*, Oxford.

FLEMING, J. A., 1939: *Terrestrial Magnetism and Electricity*, New York and London.

HEWSON, E. W., 1937: 'A Survey of Facts and Theories of the Aurora', *Rev. Mod. Physics*, **1**, 403.

ALFVÉN, H., 1939: 'Theory of Magnetic Storms, I', *Kungl. Sv. Vet.-Akademiens Handl.* (3), **18**, No. 3.

—— 1940: 'Theory of Magnetic Storms, II, III', ibid. **18**, No. 9.

BARTELS, J., 1936: 'Eccentric Dipole approximating the Earth's Magnetic Field', *Terr. Mag.*, **41**, 225.

BERNARD, R., 1940: 'Recherches sur les conditions d'excitation des divers systèmes de bandes de la molécule d'azote neutre et ionisée', *Ann. d. Physique* (11), **13**, 5.

BIRKELAND, K., 1908, 1913: *Norwegian Aurora polaris Exp. 1902–1903*, Christiania (Oslo).

Carlheim-Gyllensköld, 1886: *Expl. int. polaires 1882–83.* Exp. Suédoise. Aurore boréale, Stockholm.

Chapman, S., 1919: 'An Outline of a Theory of Magnetic Storms', *Proc. Roy. Soc.* A, **95**, 61.

—— 1927: 'On Certain Characteristics of World Wide Magnetic Disturbances', ibid. A, **115**, 242.

—— 1936: *The Earth's Magnetism.* London, Methuen's monograph.

—— and Bartels, J., 1940: *Geomagnetism,* Oxford.

—— and Ferraro, V. C. A., 1931, 1932: 'A New Theory of Magnetic Storms', *Terr. Mag.*, **36**, 77, 171; **37**, 147.

Cowling, T. G., 1942: 'On Alfvén's Theory of Magnetic Storms and of the Aurora', ibid. **47**, 209.

Harang, L., 1937: *Int. Union of Geod. and Geoph. Bull.*, **10**, 120, Copenhagen.

Hewson, E. W., 1937: 'A Survey of Facts and Theories of the Aurora', *Rev. Mod. Phys.*, **9**, 403.

Internationale Polarforschung 1882–3. *Beob. d. deutschen Stat.* **1**, Kingua Fjord, Berlin, 1886.

Malmfors, K. G., 1946: 'Experiments on the Aurorae', *Ark. f. mat., astr. o. fysik,* **34 B**, No. 1.

Paulsen, A., 1893: *Obs. int. polaires 1882–83.* Exp. Danoise, Copenhague.

Penning, F. M., 1937: 'Ein neues Manometer für niedrige Gasdrucke, insbesondere zwischen 10^{-3} und 10^{-5} mm.', *Physica,* **4**, 71.

Tromholt, S., 1882: *Sur les périodes de l'aurore boréale,* Copenhague.

Vegard, L., 1939: *Terrestrial Magnetism and Electricity,* New York and London.

—— and Kvifte, G., 1945: 'Spectral Investigations of Aurorae and Twilight', *Geofys. Publ. Oslo,* **16**, No. 7.

Vestine, E. H., 1938: 'Asymmetrical Characteristics of the Earth's Magnetic Disturbance-Field', *Terr. Mag.*, **43**, 261.

—— and Chapman, S., 1938: 'The Electric Current-System of Geomagnetic Disturbance', ibid. **43**, 366.

VII

COSMIC RADIATION

7.1. Cosmic radiation ('C.R.') has proved to be of great importance in several branches of physics. In fact it has three different aspects: the nuclear physical, the geophysical, and the cosmic physical. In this connexion we are of course mainly interested in the third of these: What does C.R. tell us about the electromagnetic conditions in space?

The answer to this question seems to be that at present the observational data are too scarce to allow any definite conclusions. Hence a discussion of the field must necessarily be of a rather speculative character, but as it brings up a series of interesting problems, we shall devote the last chapter to it.

When C.R. hits the top of the atmosphere it consists mainly of charged particles with enormous energies. There is no definite evidence that the primary radiation contains any neutral particles such as neutrons or γ-quanta. The east–west effect indicates that there are probably more positive than negative incident particles. According to Johnson (1939) and Schein, Jesse, and Wollan (1941) the primary radiation may consist of protons only, but many authors maintain that it also contains electrons. Recently Hulsizer and Rossi (1948) report that the number of primary electrons is certainly less than 1 per cent. It is possible that the protons are mixed up with other atomic nuclei.

As the terrestrial magnetic field permits particles of a certain momentum to reach the upper atmosphere of the earth only at latitudes higher than a certain critical latitude (as shown by Störmer), the intensity of C.R. is a function of the latitude. Investigations of this latitude effect make it possible to determine the momentum spectrum of the primary rays up to a few times 10^{10} e.v./c,† above which limit rays can reach the earth even at the equator. The result is that most incident particles have momenta of the order of 10^{10} e.v./c.

Measurements in Wilson chambers in strong magnetic fields have demonstrated the existence of particles with momenta up to between 10^{10} and 10^{11} e.v./c. No definite proofs of higher momenta are at present available, but the extensive showers (Auger showers) indicate

† The momentum p of a C.R. particle is conveniently expressed in electron volts divided by velocity of light. For negligible rest-mass the energy equals the momentum multiplied by velocity of light.

the existence of very much higher values. In fact, if one primary particle has been the carrier of all the energy developed in an extensive shower, it must have possessed a momentum which in some cases exceeded 10^{15} or even 10^{16} e.v./c.

The total energy which the earth receives as C.R. is about the same as that received as starlight.

The primary C.R. has a remarkable constancy and isotropy. Most of the variations in intensity registered at a certain observatory can be traced back to variations of atmospheric pressure and temperature. After correction for the atmospheric absorption there remain variations (up to about 5 per cent., in extreme cases 10 per cent.) associated with magnetic storms. No other aperiodical variations have been established with certainty. As the magnetic storm variations are produced near the earth or in any case within the solar system, the radiation in interstellar space seems to have remained constant, within the accuracy of measurements, during the couple of decades during which measurements have been made.

As the earth rotates, an observer on the earth receives radiation from different parts of the sky. Hence the sidereal time variation is a measure of the anisotropy of C.R. No sidereal time variation in excess of a few tenths of one per cent. has been established. As thc terrestrial magnetic field smoothes the diurnal variation to some extent, the isotropy need not be as good as that, but there is certainly no excess of radiation in any direction exceeding a few per cent.

It is beyond the scope of this book to treat the behaviour of C.R. in the atmosphere or to give a detailed analysis of the complicated behaviour of the rays in the terrestrial magnetic field (see Heisenberg, 1943, and Jánossy, 1948). Instead we shall confine ourselves to an analysis of the conditions before the radiation reaches the earth. Of the phenomena occurring, say, within the boundaries of the solar system the *influence of the solar magnetic field* and the *magnetic storm variations* are the most important. After a discussion of them we shall approach the fascinating problems of how the cosmic rays have acquired their enormous energy and why the radiation is isotropic.

7.2. Cosmic radiation in the terrestrial and solar magnetic fields

Consider C.R. of a certain energy outside the terrestrial and solar magnetic fields. This radiation is probably isotropic, which means that an observer measures the same intensity from all directions. When an isotropic radiation enters a magnetic field this field may prohibit the

radiation from reaching a certain point from certain directions, but in those directions which are still allowed, the intensity is unaffected by the magnetic field, as shown by Lemaître and Vallarta (1933). In other words, whether our observer makes his measurements in the solar or terrestrial field, he still finds the same intensity (of the chosen energy) as in the absence of a magnetic field when looking in some directions (allowed cone), but may find the intensity zero in other directions (forbidden cone).

7.21. *Terrestrial field.* Treating the terrestrial magnetic field as approximately a dipole field, and neglecting the solar magnetic field we find according to § 2.4 that the direction of a ray reaching the magnetic equator is given by equation 2.4 (13):

$$\sin\theta = -\frac{c_{\mathrm{St}}}{R_0}\left(2\gamma+\frac{c_{\mathrm{St}}}{R_0}\right). \tag{1}$$

Here θ is the angle which the path makes with the tangent to the equator. The parameter c_{St} is inversely proportional to the square root of the momentum of the particle [according to 2.4 (5)] and R_0 represents the radius of earth.

If we want to know whether rays of a given momentum are allowed to reach a given point from a certain direction, we must study the path arriving from that direction. If this path has cut the earth's surface, it is obvious that no particles can arrive along it. Hence it belongs to the forbidden cone. On the other hand, if it comes directly from infinity it belongs to the allowed cone and gives the full intensity of C.R. This shows that the forbidden cone may be considered as the shadow of the earth. All periodic orbits are forbidden, because a periodic orbit which reaches an observer on the earth's surface must earlier have intersected the same surface.

At the equator, rays from all directions above the horizon are allowed if, for $\gamma = -1$, $\sin\theta = +1$, which according to 2.4 (14) corresponds to the momentum†

$$p_2 = \frac{ea}{cR_0^2}, \tag{2}$$

whereas $\sin\theta = -1$ (all directions forbidden) corresponds to

$$p_1 = \frac{ea}{cR_0^2}(3-2\sqrt{2}). \tag{3}$$

Rays with momentum above p_2 reach the equator from all directions, rays below p_1 do not reach it at all, whereas rays in the intermediate

† If p is expressed in e.v./c we have $p_2 = 300aR_0^{-2}$; $p_1 = 300(3-2\sqrt{2})aR_0^{-2}$.

range are allowed within a circular cone, positive particles arriving from the west, negative from the east.

For the earth we have ($a = 8{\cdot}1\,.\,10^{25}$; $R_0 = 6{\cdot}37\,.\,10^{8}$ cm.)

$$p_1 = 1{\cdot}0\,.\,10^{10}\ \text{e.v.}/c;\quad p_2 = 6\,.\,10^{10}\ \text{e.v.}/c. \tag{4}$$

If we go to higher latitudes, p_1 and p_2 decrease so that also softer radiation reaches the earth. The shape of the allowed cones becomes very complicated. The observed effect upon C.R. of the terrestrial magnetic field is that the measured total intensity increases with latitude from the equator up to about 45°. The increase is very much larger in the upper atmosphere, where even the low-energy part of C.R. is measured, than at sea-level, to which only the most energetic particles can penetrate. In the upper atmosphere the intensity increases at least up to 58°, as shown by Millikan, Neher, and Pickering (1943). This shows that there are primary particles with momenta as low as $2\,.\,10^{9}$ e.v./c.

7.22. *Solar field.* Jánossy (1937) pointed out that the solar magnetic field may prohibit the arrival on the earth of low-energy particles. The problem has been discussed further by Vallarta (1937) and by Epstein (1938). If the earth is supposed to be situated close to the solar magnetic equatorial plane, the problem of the motion of C.R. in the solar field is the same as that which we have just discussed. We may apply equations (1), (2), and (3) if a now stands for the solar dipole moment and R_0 for the distance sun–earth. Positive particles in orbits from infinity in the range $P_1 < P < P_2$ reach the earth from the evening side, i.e. if the earth had no magnetic field they would be observed as coming from the zenith at that point of the equator which has local time 18^{h} (inclination of the earth's axis neglected). For the earth's distance ($R = 1{\cdot}5\,.\,10^{13}$ cm.) and the adopted value of the solar magnetic moment ($a = 4{\cdot}2\,.\,10^{33}$ gauss cm.^{3}) we have

$$P_1 = 1{\cdot}0\,.\,10^{9};\quad P_2 = 5{\cdot}6\,.\,10^{9}\ \text{e.v.}/c. \tag{5}$$

The problem of the motion of particles in the solar magnetic field differs from the analogous problem in the terrestrial field in one respect. In the latter case all periodic and quasi-periodic orbits reaching the earth's surface are forbidden because they are 'shadowed' by the earth itself, but a similar orbit in the solar magnetic field, which reaches the neighbourhood of the earth, has in general not been intercepted. In fact, if we trace such an orbit backwards, the chance that we shall find it to have cut the earth or any other planet earlier within a reasonable time is very small. Of the periodic and quasi-periodic orbits those resembling the orbit of Fig. 2.5 are most important.

It is obvious that all orbits from infinity must have full intensity. It has usually been assumed that the periodic orbits in the solar magnetic field should have intensity zero, because no particles from infinity can reach them. This means that the earth should receive a directed radiation in the range P_1 to P_2, so that the solar magnetic field would give rise to a solar diurnal variation.

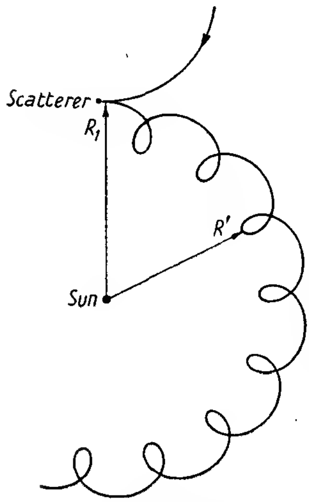

FIG. 7.1. Orbits in solar equatorial plane.

This conclusion is probably not legitimate, because we have neglected the scattering of C.R. The scattering may be due to different causes. One of them is that in the regular solar magnetic field the fields of the planets introduce irregularities. A C.R. particle which accidentally passes near a planet will be deviated ('scattered') from its regular path. Suppose that a C.R. particle enters the terrestrial field but does not hit the earth. Then it will leave the field again but deflected through a certain angle. In order to be scattered considerably a particle must approach so near to the earth that its radius of curvature ρ in the terrestrial magnetic field H is of the same order as the distance R to the earth's magnetic dipole (moment $= a_{\oplus}$). If P is the momentum of the particle we have

$$R \approx \rho = \frac{P}{H} \approx \frac{PR^3}{a_{\oplus}}, \tag{6}$$

or

$$R^2 \approx \frac{a_{\oplus}}{P}. \tag{7}$$

For $P = 10^7$ gauss cm. $= 3.10^9$ e.v./c [compare (5)], and with $a_{\oplus} = 8.10^{25}$ gauss cm.^{3}, we find for the scattering cross-section of the terrestrial magnetic field

$$S = \pi R^2 = 2{\cdot}5.10^{19}\ \text{cm.}^2 \tag{8}$$

The earth's 'cross-section for absorption' is much smaller at this energy.

Hence the effect of the terrestrial field is roughly that particles arriving within this surface will be scattered at random from one orbit in the solar field to another. Thus if a particle in an orbit from infinity approaches the earth, it may be scattered into a periodic orbit and vice versa.

When a particle has been scattered into a periodic orbit, it remains

there until it becomes absorbed, e.g. in interplanetary matter, or scattered back into an orbit to infinity by the terrestrial magnetic field. The probability that a particle in a periodic orbit is scattered by the terrestrial field may be estimated in the following way. In the solar field the periodic orbits passing near the earth occupy a volume V, $V \approx R_{\oplus}^3$, where $R_{\oplus}$ means the earth's orbital radius. A particle moving with velocity c within this volume has a fair chance of hitting a surface S after a time T,

$$T = V/Sc. \tag{9}$$

If S means the scattering cross-section of the earth's magnetic field, we find

$$T = 160 \text{ years}. \tag{9'}$$

Consequently a particle in a periodic orbit is likely to remain there one or two centuries before it is released by a new scattering.

To take an optical analogy, the solar magnetic field constitutes a 'screen' between the different types of orbits. The presence of the scatterer means a 'hole' in this screen through which particles may pass from one side to the other. The isotropic radiation, reaching the scatterer from infinity, will leak through the hole and fill also all periodic orbits passing the scatterer. A stationary state is reached when the intensity inside the screen is the same as outside it. Then the same quantity of radiation passes the hole in either direction.

If the particles in periodic orbits are absorbed considerably during the time T, the number of particles in periodic orbits is determined by the absorption in interplanetary matter during this time. If the density is ρ g. cm.$^{-3}$, the matter which the radiation passes in the time T is given by $D = cT\rho = 1{\cdot}5.10^{20}\rho$ g. cm.$^{-2}$ According to van de Hulst (1947) the density of interplanetary matter is $\rho \sim 5.10^{-21}$ g. cm.$^{-3}$ This gives $D = 0{\cdot}75$ g. cm.$^{-2}$, which means that the absorption is small.

Consequently the periodic orbits are filled to about the same intensity as the orbits from infinity. Hence for momenta above P_1 cosmic rays are reaching the earth from all directions. Below P_1 all directions are forbidden unless scattering by the outer planets or other causes makes some weaker radiation leak in.

Hence theoretically the solar magnetic field is not likely to produce a diurnal variation. This agrees with earlier results by Malmfors (1945), who by a study of the trajectories has shown that the observed solar time variation at Stockholm (magn. lat. = 58°) cannot be a product of the solar magnetic field.

It is probable that all planets have magnetic fields. Whether we

suppose these fields to be proportional to the angular momenta or, for example, to the volumes of the bodies, we find that all planets, except the asteroids, are likely to be rather effective scatterers. It can be shown that the earth should be reached by radiation scattered by a magnetic field of Mars. This effect makes the low momentum limit of C.R. about 30 per cent. less than that given by equation (5). Other planets are of no importance in this respect.†

Cosmic dust and free atoms, ions, and electrons may also act as scatterers, but it is unlikely that they are of any importance.

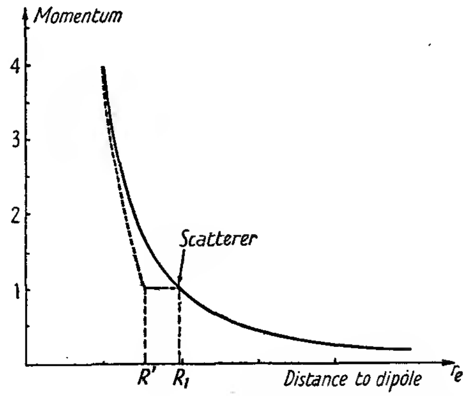

FIG. 7.2. Minimum momentum with and without scatterer.

More effective are electric fields. In §§ 5.7 and 6.1 we have seen that in connexion with magnetic storms the sun emits beams, which are associated with electric fields. The deviation of a C.R. particle when passing once through such a field may not be very large, but by repeated scattering the effect may accumulate. As found above, the absorption is so small that a C.R. particle may move for centuries in a periodic orbit in the solar field. It seems likely that all sorts of disturbances would change the orbit completely during so long a time. If we accept this, the conclusion should be that there exist no forbidden orbits in the solar magnetic field. C.R. of all energies present in interstellar space fill up interplanetary space. Only close to the solar surface are there forbidden regions which are due to the 'shadow' of sun itself (as in the case of the earth).

† Since this was written Kane, Shanley, and Wheeler (*Rev. Mod. Phys.*, 1949) have made a very careful study of the motion of C.R. in the solar magnetic field, taking account of the scattering and absorption by the planets and the absorption by the moon and the sun.

7.3. Magnetic storm variations

Large variations in C.R. intensity, which cannot be accounted for by changes in atmospheric absorption, are observed only in connexion with magnetic storms. Fig. 7.3 shows records of C.R. intensity at different

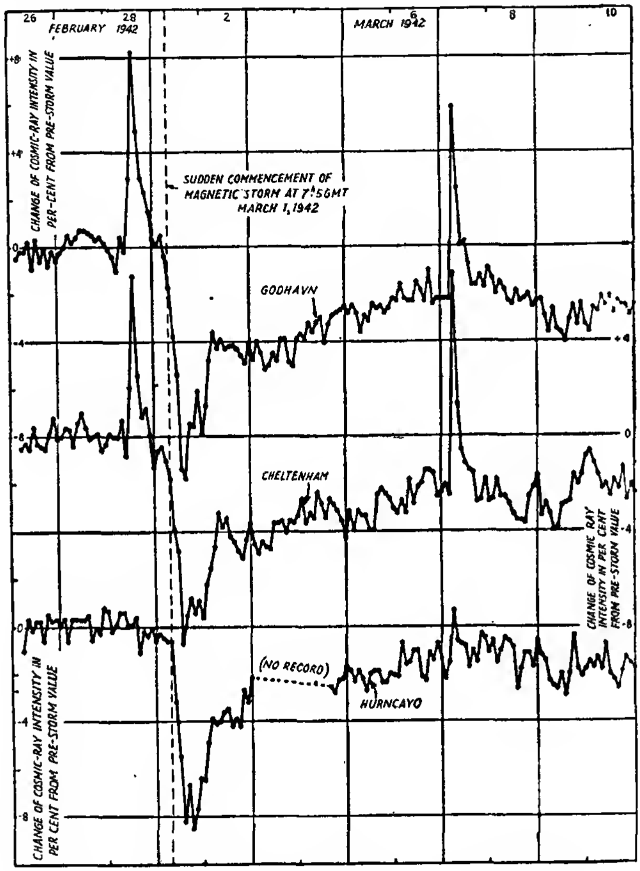

FIG. 7.3. C.R. variation during a magnetic storm. (Lange–Forbush, 1942).

observatories during two magnetic storms. From these and other measurements it is evident that these variations occur simultaneously on all stations. The variations are associated with magnetic storms, but there is no simple connexion. In general C.R. intensity increases some hours before the magnetic effects commence and decreases during the main phase of the storm.

For that part of C.R. which is hard enough to penetrate the atmosphere there are no forbidden cones to an observer at a latitude above about 45°

(see § 7.21). A change of the terrestrial magnetic field may introduce forbidden cones at higher latitudes and thus cause a *decrease* in C.R., but it cannot possibly cause an *increase*, because the full intensity is already normally observed. The fact that C.R. increases at Godhavn (geomagnetic latitude $= 80°$) during some part of a magnetic storm shows that changes in the earth's magnetic field cannot be the (direct) cause of the C.R. storm variation.

Nor can the solar magnetic field cause the observed C.R. variations. This is shown by the fact that even at an equatorial station such as Huancayo (geom. lat. $= 0{\cdot}6°$) variations are observed. At the equator only the most energetic particles (momentum $> 10^{10}$ e.v./c) are permitted, and for these there are no forbidden cones in the *solar* magnetic field, which according to 7.22 (5) only affects particles with momenta below $5{\cdot}6.10^9$ e.v./c.

Hence the only possible explanation seems to be that the storm variations of C.R. are due to changes in the electric conditions around the earth. According to 5.81 (10) the electric field in interplanetary space is

$$\mathbf{E} = -\frac{\mu}{c}[\mathbf{vH}]. \tag{1}$$

In connexion with magnetic storms a stream is sent out from the sun with a high velocity. The solar magnetic field polarizes the stream and this polarization is compensated by the electric field (1) in order to make the current zero within the stream. If a C.R. particle passes the stream, it is acted upon by the electric field E (but not by the polarization, which is effective only for the particles moving with the stream). Hence if a C.R. particle crosses the stream its energy changes by

$$\Delta W = e\int E\,ds. \tag{2}$$

Outside the stream there is no electric field (in the ideal case; compare § 5.8). It should be observed that the electric field is not derivable from a potential.

Therefore, in order to see what happens to C.R. when it passes a stream, let us consider an electric double layer giving a change in voltage of ΔV $(= \int E\,ds)$. On one side of the double layer we have the normal isotropic C.R. What is observed on the other side of the layer?

A positive particle with the energy W, making the angle φ with the normal to the layer, will emerge at the other side with the energy

$$W' = W+\Delta W$$

and at the angle φ', which if rest mass is negligible in comparison with kinetic mass is given by

$$\frac{\sin\varphi'}{\sin\varphi} = \frac{W}{W+\Delta W}. \tag{3}$$

This can easily be obtained by elementary calculations or from the fact that the refractive index is inversely proportional to the length of the de Broglie waves, which is inversely proportional to the momentum, which at relativistic velocities is proportional to the energy.

As the intensity (density of the particles) is inversely proportional to the solid angle covered, the intensity increases by the factor

$$(W+\Delta W)/W.$$

Consequently, after the passage of the double layer, positive particles with the energy W will increase in number per solid angle by a factor $(W+\Delta W)/W$ and, at the same time, their energy will increase to $W+\Delta W$. For negative particles there is a corresponding decrease in number and energy.

Confining ourselves to positive primaries (which in any case produce most of the effects observed at sea-level) we suppose that the number of primaries in the energy interval W to $W+dW$ is $n(W)\,dW$ and that one primary particle of energy W after passing the atmosphere gives a reading $\alpha(W)$ on a measuring apparatus. Then the whole measured intensity is

$$I = \int_0^\infty \alpha(W)n(W)\,dW. \tag{4}$$

If all particles gain the energy ΔW in the double layer, the intensity becomes

$$I+\Delta I = \int_0^\infty \left(1+\frac{\Delta W}{W}\right)\alpha(W+\Delta W)n(W)\,dW.$$

For small values of ΔW we have

$$I+\Delta I = \int_0^\infty \left(1+\frac{\Delta W}{W}+\frac{\Delta W}{\alpha}\frac{d\alpha}{dW}\right)\alpha(W)n(W)\,dW, \tag{5}$$

or

$$\Delta I = \Delta W \int_0^\infty \alpha n\left(\frac{1}{W}+\frac{1}{\alpha}\frac{d\alpha}{dW}\right)dW. \tag{6}$$

The integral in (6) is difficult to evaluate because α is not known. The relative increase in intensity $\Delta I/I$ is probably about equal to the relative increase in energy $\Delta W/W_0$ of the rays near the maximum in the spectrum ($W_0 \approx 10^{10}$ e.v.), according to a crude analysis.

Chapman has assumed that a storm-producing stream moves with a constant radial velocity, the tangential motion being governed by the constancy of the angular momentum (which equals the momentum at the solar surface). Fig. 7.4 shows the shape of a stream at a certain instant. Although electromagnetic forces may act upon the stream (e.g. they tend to establish isorotation), it seems reasonable to start with Chapman's assumption.

The position of the earth before the stream has arrived and after it has passed is shown in the figure. The field of the stream is directed forward. Hence before the stream has reached the earth positive particles which

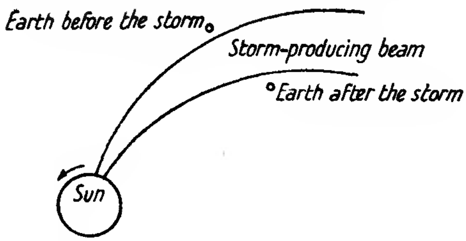

FIG. 7.4. Shape of stream from the sun according to Chapman.

pass through it on their way to the earth are accelerated, so that an increase in C.R. intensity is caused. After the stream has passed the earth, positive particles directed towards the earth lose energy when crossing the stream, so that the C.R. intensity diminishes.

Consequently the storm variations in C.R. are explained qualitatively. That the earth before the storm is situated on the convex side and after the storm on the concave side of the stream also explains why the increase before the storm is of shorter duration than the decrease after the storm (when the earth is partially wrapped up by the stream).

The quantitative side is more difficult. In order to produce an intensity change of as much as 10 per cent., ΔV cannot be much smaller than about 10^9 volts. The breadth of the stream near the earth cannot exceed about 10^{13} cm., so that the field must not be much less than 100 μ volt/cm. A velocity v equal to 3.10^8 cm. sec.$^{-1}$ in a field H equal to $1{\cdot}2.10^{-6}$ gauss (solar field at the earth's distance) produces only 4 μ volt/cm. This is probably enough for ordinary storm variations in C.R., which amount to only 1 per cent. or less. A 10 per cent. variation is certainly an extraordinary event, but still it must be explained. The solar magnetic field is not likely to be in error by more than a factor 2, so we should have to assume a velocity of the order 3.10^9 cm./sec. This would mean enormous

kinetic energy of the ions in the stream. Perhaps it would be possible to assume that the stream, which comes from the solar surface where the field is of the order of 10 gauss, in part carries a strong magnetic field with itself. Owing to the high conductivity the field may be 'frozen in'. A careful analysis of the different possibilities is needed.

If cosmic rays move in periodic orbits in the solar field (see § 7.22), these are affected very much by a storm-producing stream, because they cross it once for every revolution around the sun. The energy of periodic orbits near the earth is small however ($< 5{\cdot}6 \,.\, 10^9$ e.v./c), so that only the weakest part of C.R. is affected in this way. Certain periodic orbits close to the sun possess higher energies (up to 10^{13} e.v./c), and if particles in these orbits are accelerated there is a chance that the periodic orbit is changed into an orbit to infinity, so that the particles leave the close vicinity of the sun and eventually reach the earth. Hence the energy change of particles in periodic orbits may produce an increase, but no decrease, in C.R. intensity measured on the earth.

Finally, it should be mentioned that if there are some C.R. particles near the solar surface, they may be accelerated by eruptive prominences. Such prominences are sometimes emitted at the speed of 10^8 cm./sec. Suppose that they move in a field $H \approx 1$ gauss and that their linear dimension is of the order $l = 10^{10}$ cm. The voltage difference over such a prominence is $V = vHl/c = 0{\cdot}33 \,.\, 10^8$ e.s.u. $= 10^{10}$ volts. Acceleration processes are discussed further in § 7.6.

7.4. The isotropy of C.R.

The smallness of the sidereal time variation indicates that C.R. in interstellar space near our solar system has a high degree of isotropy. If C.R. consisted of γ-quanta, as was initially assumed, or other neutral particles, which necessarily travel rectilinearly through space, the isotropy at our place must mean, unless very special assumptions are made, that the radiation is isotropic everywhere in the universe, even outside our galaxy. Near our solar system the energy density of C.R. is about the same as that of starlight (sun excluded). As in intergalactic space the density of starlight is several powers of 10 less than in our neighbourhood, the total energy of C.R. in the universe must be several powers of 10 times as great as the total starlight energy. Hence the 'universal isotropy' of C.R. means that C.R. plays a decisive role in the energy balance of the universe. The source producing cosmic rays is far more energetic than all the stars together.

The discovery that C.R. consists of charged particles makes it possible

to avoid the above conclusion, because the paths of charged particles may be curved by magnetic fields. Electric fields of reasonable strength are unimportant. Consider the case when in an arbitrary magnetic field, say a dipole field, there are situated a number of sources emitting charged particles with such small momenta that the radii of curvature of their paths in the magnetic field are small in comparison with the extension of the field. The particles will oscillate along the lines of force and also drift because of the inhomogeneity of the field. An observer situated not too close to any of the sources will have a good chance to receive charged particles from all directions, i.e. to observe a radiation with a high degree of isotropy.

If sources within our galaxy are to produce a radiation which is fairly isotropic within the galaxy, we require a galactic magnetic field H strong enough to bend the particle paths so that the radius of curvature ρ becomes much smaller than the extension of the galactic magnetic field

$$\rho = \frac{P_v}{300H}, \tag{1}$$

where P_v = momentum in e.v./c. As our galaxy is a flat disk with a diameter of 100,000 light years (10^{23} cm.) and a thickness of 10,000 light years (10^{22} cm.) we must require $\rho \leqslant 10^{22}$ cm. This gives for $P_v = 3.10^9$ e.v./c, $H \geqslant 10^{-15}$ gauss; for $P_v = 3.10^{12}$ e.v./c, $H \geqslant 10^{-12}$ gauss; and for $P_v = 3.10^{15}$ e.v./c, $H \geqslant 10^{-9}$ gauss. Consequently, as the main part of C.R. has momenta below, say, 3.10^{12} volts, a galactic magnetic field of 10^{-12} gauss is enough to make it fairly isotropic, but in order to make even the radiation supposed to produce the extensive showers isotropic, which it seems to be, a field of 10^{-9} gauss is required.

(The magnetic fields of the stars are of no importance in this connexion. At a distance of one light year from the sun the solar field is only 10^{-22} gauss and the average field from the stars cannot be more than a few orders of magnitude greater.)

If we assume the existence of a galactic magnetic field, it is possible that the intensity of C.R. is much smaller outside the galaxy than within it. This would be the case if the sources of C.R. are situated inside the galaxy, e.g. near the stars. The galactic magnetic field may very well have such a shape that only a small fraction of the C.R. of the galaxy leaks out into intergalactic space.

Under these conditions the output of the sources of C.R. may be much smaller than the energy dissipation of the stars. If we suppose that the leakage from the galactic magnetic field is negligible, a stationary state

is reached when the generation of cosmic rays equals the absorption. The main absorption will take place in interstellar matter. The average density of this is usually estimated to be about 10^{-24} g. cm.$^{-3}$ If a C.R. particle can penetrate say 100 g./cm.^{2}, its range in interstellar space is 10^{26} cm. or 10^8 light years. As the energy density of C.R. and starlight is about the same but the average path of starlight within the galaxy is only 10^4 light years, the production of C.R. need only be 10^{-4} of the starlight production.

Consequently the fact that C.R. consists of charged particles gives us two possible ways of interpreting the isotropy:

1. '*Universal isotropy hypothesis.*' The cosmic radiation is isotropic in the whole universe. The galactic magnetic field may have any value (including zero). The total energy of C.R. in the universe is several powers of 10 larger than the starlight energy.

2. '*Galactic isotropy hypothesis.*' The C.R. is 'enclosed' in the galactic system, where the C.R particles spiral around the lines of force of a magnetic field which is at least 10^{-12}–10^{-9} gauss. Outside the galaxy the intensity may be much smaller than within it. The total energy of C.R. in the universe may be only a small fraction ($\approx 10^{-4}$) of the starlight energy.

7.5. Speculations about a galactic magnetic field

There is no direct evidence of the existence of a galactic magnetic field, but the assumption of such a field makes it easier to explain the origin and isotropy of C.R. In any case it is of interest to try to analyse the possible properties of the assumed field. According to § 7.4 the strength should be at least of the order 10^{-9}–10^{-12} gauss.

The solar magnetic field, supposed to be a dipole field, is of the order 10^{-10} gauss at Pluto's orbit, so it is of importance only within the limits of the planetary system. Even if other stars have somewhat stronger fields, all the star fields are too local to be of importance in this connexion. The galactic magnetic field must be, so to say, a property of interstellar space, or rather of interstellar matter.

The density of interstellar matter is estimated to be 10^{-24} g./cm.^{3} or about one atom per cm.^{3} This is certainly a very low density, but as the mean free path (of the order of 10^{15} cm.) is much smaller than the dimensions of the galaxy (10^{22}–10^{23} cm.) and the time between two consecutive collisions of a molecule (of the order of 10^8 sec.) is small in comparison with the age of the universe, there is no reason why we should not treat it as an ordinary gas. Its temperature is supposed to be 10,000°. Hence

according to § 3.24 the electrical conductivity parallel to the magnetic field $\sigma_{\parallel} = 1{\cdot}4.10^{13}$ e.s.u. Its conductivity perpendicular to the field, $\sigma_{\perp}$, is for $H = 10^{-9}$ gauss equal to $1{\cdot}4.10^{7}$ e.s.u. and for $H = 10^{-12}$ gauss $1{\cdot}4.10^{13}$ e.s.u. Consequently interstellar space should be regarded as a good but anisotropic conductor. In fact the parallel conductivity is of the same order as in the sun. According to § 3.12 the conductivity is transformed as η^{-1}, so when reduced to 'laboratory scale' the galaxy gets a very high conductivity.

The decay constant T of the sun's general magnetic field is according to Cowling (1945) of the order 10^{10} years (see § 5.22). As T is proportional to l^2 (where l is the linear dimension) we find that the galaxy, the radius of which is 10^{12} times the solar radius, ought to have a decay constant of 10^{34} years. This shows that the magnetic lines of force are very effectively 'frozen' into the interstellar matter. No appreciable change of the magnetic field can occur during the age of the universe, unless in connexion with the motion of interstellar matter.

With $H \leqslant 10^{-9}$ gauss the velocity of magneto-hydrodynamic waves $V = H(4\pi\rho)^{-\frac{1}{2}} \leqslant 300$ cm. sec.$^{-1} = 10^{-8}c$. Hence during the age of the universe (10^{10} years) a wave travels only 100 light years. Because of the low wave velocity the magnetic field does not affect the motion of interstellar matter very much. The magnetic pressure ($H^2/8\pi \leqslant 4.10^{-20}$ dyn. cm.$^{-2}$) is much lower than the gas pressure ($1{\cdot}4.10^{-12}$ dyn. cm.$^{-2}$). Consequently, unless the galactic magnetic field is much stronger than assumed, its effect upon interstellar matter is negligible.

The long decay constant makes it futile to look for present causes of the galactic magnetic field. It has been assumed (by the author) that a small anisotropy of C.R., which is equivalent to a current through space, could cause a galactic magnetic field. This is not right. Suppose that a beam (of cosmic dimensions!) of, say, swift electrons is emitted. The current of the beam tends to produce a magnetic field, but because of the high conductivity no appreciable change in the magnetic state is permitted. The result is instead that a compensating current is induced which carries slow interstellar electrons in a direction opposite to the electrons of the beam, so that the sum of the currents becomes very close to zero.

The long decay constant shows that the galactic magnetic field is probably a relic from a primeval state.

Blackett (1947) has pointed out that a magnetic field may be a general property of all rotating bodies (compare § 1.1). He finds for the earth, sun, and 78 Virginis that the magnetic dipole moment a is proportional

to the angular momentum U:

$$a \approx 10^{-15} U. \tag{1}$$

Let M be the mass, R the radius, and v the average velocity of the galaxy due to rotation. Then we have (for the order of magnitude) $U = MRv$ and

$$H = aR^{-3} = 10^{-15} MvR^{-2}. \tag{2}$$

The total mass is supposed to be 10^{11} times the solar mass, or about 10^{45} g. With $v = 10^7$ cm. sec.$^{-1}$ and $R = 3.10^{22}$ cm. we obtain $H = 10^{-8}$ gauss. If instead we take only the mass of interstellar matter, supposed to have a uniform density ρ ($= 10^{-24}$ g. cm.$^{-3}$), we have $H = 10^{-15} dv$ (where d = thickness of the galactic disk). With $d = 10^{22}$ cm., we obtain $H = 10^{-10}$ gauss. The field according to Blackett's relation is about what is required to make even the extensive shower particles fairly isotropic. It must be pointed out that if there is a general connexion between magnetic field and rotation, as Blackett assumes, this means that a new term must be introduced into Maxwell's equations, a term which certainly is very important in cosmic electrodynamics.

Let us assume that once upon a time, say 10^{10} years ago, the galactic magnetic field had a simple shape, e.g. resembled the sun's general magnetic field. In the solar as well as in the galactic field the lines of force are 'frozen' into the matter so that a local motion of matter drags the lines of force with it. This causes magneto-hydrodynamic waves which upon the whole try to re-establish the primary state. Any local motion is braked in a time which equals the time a wave needs to travel over the region which was put into motion. As a wave travels through the whole sun in less than 100 years, the waves succeed upon the whole in keeping the solar magnetic field 'in order'. This is not the case for the galactic magnetic field. If a region of an extension of, say, 1,000 light years starts moving, it will not be braked by magneto-hydrodynamic phenomena until after more than 10^{11} years (on the above assumptions $H \approx 10^{-9}$ gauss) which is more than the age of the universe. Hence, as the restoration is very slow, local motions will deform the magnetic field and drag out the lines of force more and more. Consequently if there is a galactic field of less than 10^{-9} gauss it is probable that it is very irregular.

With increasing disorder the average magnetic field will increase because the lines of force are dragged out more and more. The average primeval field may have been much weaker than the present one.

If the field were as strong as 10^{-6} gauss, a wave could pass the whole galaxy in a time less than the age of the universe. Such a field would be regular and would also regulate the state of motion of interstellar matter.

7.6. Origin of C.R.

As we have seen, there are two different ways to interpret the observed isotropy of C.R. According to the 'universal isotropy hypothesis', the C.R. is really isotropic in the whole universe (or a very considerable part of it) and the total energy of the cosmic radiation is many times as large as that of starlight. If this is right, the origin of C.R. is to be found along cosmological lines. As Lemaître (1931) has pointed out, C.R. may have been generated when the conditions in the universe were quite different from those prevailing at present, the radiation circulating in the closed universe since that time. The 'pre-stellar state', which is supposed to have preceded our present 'stellar state' and during which the elements are supposed to have been formed, had a temperature of 10^{11} degrees according to investigations by Chandrasekhar and Heinrich (1942) and by Klein, Beskow, and Treffenberg (1946). Although this is extremely high in some respects, it corresponds to only 10^7 e.v. and thus does not give very much hope of explaining the energies of C.R. Thus the generation of C.R. must probably be put back to a still earlier period of which we know nothing.

Millikan has for long advocated a theory according to which the origin is due to a spontaneous annihilation of atomic nuclei, a process which, however, is extremely unlikely from a theoretical point of view. As support of his theory he quotes that he and his collaborators (Millikan, Neher, and Pickering, 1943) have found that the energy spectrum of C.R. has a band structure, but this result has not been confirmed by others.

Klein (1944) has pointed out that general symmetry considerations suggest that part of the universe consists of 'inverse matter' with negative atomic nuclei and positive electrons. Collisions between inverse nuclei and ordinary nuclei are likely to give rise to a very energetic (10^9–10^{10} e.v.) annihilation radiation. Although this theory is perhaps the most attractive one of the 'universal isotropy' type, it probably meets insuperable difficulties in explaining the higher parts of the energy spectrum.

On the other hand, we may assume that the observed isotropy is produced by a galactic magnetic field ('galactic isotropy hypothesis'). The assumption of a galactic magnetic field does not invalidate any theory of the type discussed above, but it gives us new possibilities of accounting for the origin of C.R., because the total amount of energy which the source must be able to produce is now very much smaller, in fact only a small fraction of the starlight energy.

The nova and supernova hypothesis—although originally proposed with the 'universal isotropy hypothesis' as background—would gain in likelihood if combined with the 'galactic isotropy hypothesis', because the total energy, generated as C.R. during these explosions, need not be so overwhelmingly large. It is difficult to see, however, how the very large particle energies of C.R. could be explained in this way. Even if the nova, and especially the supernova, explosions are very violent phenomena and associated with temperatures which from an ordinary point of view are enormously high, there is no likelihood that single particles could acquire energies of the order of magnitude found in C.R.

The possibility that C.R. has got its energy by acceleration in electrostatic fields has been pointed out by several authors, first by Bothe and Kolhörster (1929). This process encounters the difficulty that interstellar space is a good conductor. In other words, there are too many charged particles in space, and an ordinary electrostatic field in a certain region must accelerate *all* charged particles which are present to C.R. energies, which requires an enormous energy if the region is extended. In order to avoid this difficulty we must either find a mechanism by which the field only accelerates some of the particles, e.g. those which already have high energies, or introduce fields which are very intense over a rather small region.

We shall discuss these two possibilities in the following.

7.61. *Acceleration in streams of ionized matter.* As we have seen in § 5.82, the high conductivity of space precludes the existence of electrostatic fields of any importance in this connexion. Induced fields may be produced by relative motion of matter. If the velocity of a stream of matter is v and the magnetic field H, the electric field is given by

$$\mathbf{E} = -\frac{\mu}{c}[\mathbf{vH}].$$

If a charged particle moves in this field and its radius of curvature ρ is so small that E and H are approximately constant over ρ, it does not change its energy. In fact, as E is perpendicular to H it only produces a drift with the velocity v, i.e. it compels the particle to move with the stream of matter. As all motions are relative, the same result is obtained if the problem is treated in a system where the stream is at rest. Then no electric field is present and the particle just spirals around a magnetic line of force.

On the other hand, if the orbit of the particle is so large that it traverses

regions with different states of motion, a change in energy is possible. An example of this has been given in § 7.3, in connexion with a theory of magnetic storm variations in C.R. A similar example will be discussed here.

Suppose that in Fig. 7.5 there is a homogeneous magnetic field H perpendicular to the paper. A stream of matter with velocity v and breadth b cuts the plane at an oblique angle. A particle moves in the plane of the paper in such a way that it crosses the stream. In order to do so its energy W must be so large that the radius of curvature

$$\rho\ (= W/eH)$$

equals at least $\frac{1}{2}b$. When it crosses the stream its energy increases by

$$\Delta W = eEb = ebv_{\perp} H/c.$$

As $W > \frac{1}{2}bH$ the maximum relative increase in energy

$$\frac{\Delta W}{W} < 2\frac{v_{\perp}}{c}.$$

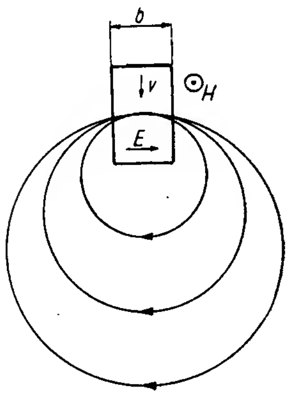

FIG. 7.5. Acceleration of particles in stream field.

The velocity of streams of matter is not likely to be close to the velocity of light. For a storm-producing stream emitted by the sun (as has been treated in §§ 5.7, 6.1, and 7.3), v/c is supposed to be about 0·01. Even if higher values are not excluded, the relative increase at one single crossing cannot exceed a few per cent. Other stars than the sun might send out corresponding streams with higher velocities, but not even this would suffice to increase the energy by orders of magnitude.

A large increase would be possible if multiple acceleration takes place. When in Fig. 7.5 the particle has crossed the stream it will move in a larger circle, which, under the ideal conditions which we have assumed, brings it back to the stream. As every crossing gives a new increase in energy, the resultant acceleration may be very large. The accelerating field is non-potential and may be interpreted as due to a change in magnetic flux produced at the upper and lower border of the stream in connexion with the magneto-hydrodynamic waves to which the stream gives rise (see §§ 4.6 and 5.82).

The acceleration process which we have discussed will take place wherever there is a stream of ionized matter. In interstellar space there are large streams in connexion with the motions of stars or star clusters.

The velocities are in general so small that the relative increase in energy does not become very important. In special cases multiple accelerations are possible, especially if particles move more or less accidentally in periodic orbits. The chance that a particle loses energy is of course about the same as that of energy gain. Hence the probability of obtaining very large energies in this way seems to be small.

More favourable are the conditions near stars. By analogy with solar conditions we could expect streams with $v/c = 0{\cdot}01$, perhaps in some cases even more. Further, there are stable periodic orbits in the dipole field of a star, so that there is a fair chance for multiple acceleration in non-potential fields of the type described. In § 7.3 the possibility of accounting for the storm increase in C.R. by multiple acceleration was mentioned. Even here the chance for energy loss is about the same as for energy gain, but if a particle in a periodic orbit has increased its energy sufficiently the orbit becomes non-periodic and the particle leaves the neighbourhood of the star. Hence of the group of particles, the energies of which change accidentally up and down, the most energetic are permanently emitted.

7.62. *Fields from stellar rotation.* As an alternative to the acceleration in fields produced by streams of ionized matter we shall also discuss the possible acceleration in fields produced at the rotation of single and double stars. At first it may be of interest to point out that if there is a general galactic field (which we have already assumed as we are working with the 'galactic hypothesis') the rotation of the galaxy may produce a polarization in the same way as, according to § 5.82, the sun is polarized. The potential is of the order $V = vHR/c$ if H is fairly regular. If H is very irregular, as was suspected in § 7.5, it becomes smaller. With $v = 10^7$ cm. sec.$^{-1}$, $H = 10^{-9}$ gauss, and $R = 3.10^{22}$ (values used in § 7.5) we find $V = 10^{10}$ e.s.u. $= 3.10^{12}$ volts. If in some way this voltage could be used to accelerate particles it would explain the energy of the main part of C.R. In order to account even for the extensive shower particles of, say, 3.10^{15} e.v. we must assume $H = 10^{-6}$, which seems not to be in conflict with any known facts. We know, however, too little, or in fact nothing, about the galactic magnetic field to be able to analyse the problem more closely, so we shall confine ourselves to stating that there may be voltage enough but it is not easy to find a mechanism by which it could be converted into particle kinetic energy.

Similar results are obtained for electric fields produced near the stars. The sun is polarized so that the voltage difference between the equator and the poles amounts to 2.10^9 volts. Stars with stronger magnetic field

and swifter rotation than the sun are able to produce larger potentials. For example, 78 Virginis has probably a magnetic field which is 30 times, an angular velocity which is 25 times, and a radius which is twice the solar values (see Babcock, 1947). This gives 6.10^{12} volts.

Swann (1933) is one of the first to point out that the acceleration could be produced in changing magnetic fields. According to him the

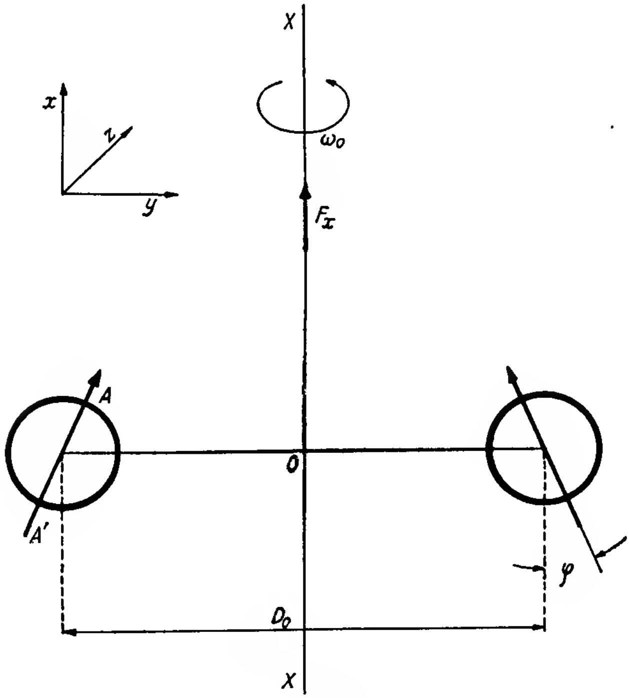

FIG. 7.6. Double-star generator.

betatron action of a rapidly increasing sunspot magnetic field may accelerate charged particles to C.R. energies.

Another model employs the electric fields induced by the rotation of a double star (Alfvén, 1937 *a*). We shall here discuss this process more closely. Let us make the following assumptions. The two components of a double star are exactly equal, both having the magnetic dipole moment a. They move along the same circle (radius $= \frac{1}{2}D_0$) with constant angular velocity ω_0 around the symmetry axis, which we take as x-axis of an orthogonal reference system (see Fig. 7.6). Their magnetic axes (which may coincide with the rotational axes of the components) are not

parallel to the x-axis but tilted in opposite directions at an angle ϕ. Their directions remain fixed during the revolution.

Because of the symmetry the magnetic field on the x-axis is always parallel (or antiparallel) to this axis. Hence charged particles can move freely along this line from $x = -\infty$ to $x = +\infty$. During this motion they may be accelerated by an induced electric field E caused by the motion of the dipoles. It is easy to calculate E_x at the instant when the two dipoles lie in the same plane (as in Fig. 7.6). One of them moves in the direction of the positive z-axis (perpendicular to the plane of the figure) with the velocity $\frac{1}{2}D_0\omega_0$. If H'_y is the y-component of the magnetic field of this dipole at a point on the x-axis, the induced electric field is $\frac{1}{2}D_0\omega_0 H'_y/c$. The other dipole moves antiparallel to the first one and its magnetic y-component has also the opposite direction. Hence the electric field deriving from the second dipole has the same magnitude and sense, so that the total electric field becomes

$$E_x = \frac{D_0\omega_0}{c} H'_y. \tag{1}$$

From the formulae for the field from a dipole [1.2 (12)–1.2 (14)], we find

$$H'_y = \frac{a}{R^5}\left(\frac{D_0^2}{2} - x^2\right)\sin\varphi + \tfrac{1}{2}D_0 x\cos\varphi, \tag{2}$$

with $R = \sqrt{\{(\frac{1}{2}D_0)^2 + x^2\}}$. The energy gained by a particle with the charge ne moving from the negative to the positive infinity along the x-axis is

$$W = neV, \tag{3}$$

with

$$V = \int_{-\infty}^{+\infty} E_x\,dx. \tag{4}$$

The integration gives

$$V = \frac{8a\omega_0}{cR_0}\sin\varphi, \tag{5}$$

where $R_0 = \frac{1}{2}D_0$. V is the electromotive force of the 'double-star generator', which accelerates positive and negative particles in opposite directions. It is easily seen that half a turn later the e.m.f. has the opposite direction. The generator gives an alternating voltage with the amplitude V. The same voltage would be induced in a conducting wire along the x-axis (the circuit being closed by a conductor at a great distance).

Suppose that D_0 is 10 times the radius of our sun and the period of rotation is 2 days (reasonable values for a narrow double star), and further that $\varphi = 10°$. Then a voltage of 10^{11} volts is obtained if a is 15

times the solar dipole moment. As the surface field of 78 Virginis is 30 times the solar value and its radius is supposed to be twice the solar radius, its dipole moment is 30.2^3 (= 240) times the solar moment. It is unlikely that 78 Virginis should represent the highest possible value among the stars. Hence it is not unreasonable to assume that in special cases a double-star generator may accelerate particles to 10^{13} e.v. or even more. If a singly ionized heavy atom is accelerated, it will soon be stripped of most of its electrons by collisions with other particles. Because of its multiple charge its final energy may be still higher by one or two powers of 10. Consequently even the highest energies observed in C.R. may be attainable by this process.

A rough estimate of the number of particles accelerated by a double-star generator and the frequency of double stars (Alfvén, 1937 *a*) seemed to indicate that the generators may be powerful enough to explain the observed intensity of C.R. (of course under the galactic hypothesis). The estimation was probably too optimistic, and the intensity problem is serious.

Another objection against Swann's theory as well as the double-star theory is that the conductivity of the solar atmosphere and of space around the stars is so high that strong fields are prohibited. In other words, the resistance in the circuit of the generator may be so low that it is practically short-circuited. Hence the result should be not a certain number of swift particles but instead a much greater number of relatively slow particles. It is difficult to judge whether this objection is right. If the current density surpasses the limit given in § 3.5, as it probably ought to do in many cases, we should expect a very high resistance and, at least at intervals, surges of very high voltage.

The result of our discussion is that there seem to be two different ways in which particles may be accelerated. The first one has the advantage of being a very general process, taking place as soon as there is a stream of ionized matter, but the disadvantage that acceleration to very high energies must occur in many successive steps. The second one has the advantage that particles are accelerated to very high energy in one single process, whereas the main objection is that the process requires somewhat special conditions.

As a general conclusion it must be stated that we cannot in any way account for C.R. without reference to cosmological problems or introduction of new physical laws. This is obviously the case if we work with the 'universal isotropy hypothesis'. If instead we chose the 'galactic hypothesis', it certainly may be possible to account for the generation

by classical processes, but we must also introduce a galactic magnetic field, and the origin of this field still remains an unsolved question.

REFERENCES

General references:

MIEHLNICKEL, E., 1938: *Höhenstrahlung*, Leipzig. (Contains a very complete bibliography.)

HEISENBERG, W., 1943: *Vorträge über kosmische Strahlung*, Berlin.

JÁNOSSY, L., 1948: *Cosmic Rays*, Oxford.

ALFVÉN, H., 1937 *a*: 'Versuch zu einer Theorie über die Entstehung der kosmischen Strahlung', *Zs. f. Phys.*, **107**, 579.

—— 1937 *b*: 'Cosmic Radiation as an Intragalactic Phenomenon', *Ark. f. mat., astr. o. fysik*, **25 B**, No. 29.

BABCOCK, H. W., 1947: 'Zeeman Effect in Stellar Spectra', *Astrophys. J.*, **105**, 105.

BLACKETT, P. M. S., 1947: 'The Magnetic Field of Massive Rotating Bodies', *Nature*, **159**, 658.

BOTHE, W., and KOLHÖRSTER, W., 1929: 'Das Wesen der Höhenstrahlung', *Zs. f. Phys.*, **56**, 751.

CHANDRASEKHAR, S., and HEINRICH, L. R., 1942: 'An Attempt to interpret the Relative Abundances of the Elements and their Isotopes', *Astrophys. J.*, **95**, 288.

COWLING, T. G., 1945: 'On the Sun's General Magnetic Field', *Mon. Not. R. Astr. Soc.*, **105**, 166.

EPSTEIN, P. S., 1938: 'Influence of the Solar Magnetic Field upon Cosmic Rays', *Phys. Rev.*, **53**, 862.

HULSIZER, R. I., and ROSSI, B., 1948: 'Search for Electrons in the Primary Cosmic Radiation', ibid. **73**, 1402.

VAN DE HULST, H. C., 1947: 'Zodiacal Light in the Solar Corona', *Astrophys. J.*, **105**, 471.

JÁNOSSY, L., 1937: 'Über einen möglichen Einfluss des Magnetfeldes der Sonne auf die in Erdnähe beobachtete Höhenstrahlung, *Zs. f. Phys.*, **104**, 430.

—— 1948: *Cosmic Rays*, Oxford.

JOHNSON, T. H., 1939: 'Evidence that Protons are the Primary Particles of the Hard Component', *Rev. Mod. Phys.*, **11**, 208.

KLEIN, O., 1944: 'On the Origin of Cosmic Radiation', *Ark. f. mat., astr. o. fysik*, **31 A**, No. 14.

——, BESKOW, G., and TREFFENBERG, L., 1946: 'On the Origin of the Abundance Distribution of Chemical Elements', ibid. **33**, No. 1.

LANGE, I., and FORBUSH, S. E., 1942: 'Effect on Cosmic-ray Intensity of the Magnetic Storm of March 1, 1942', *Terr. Mag.*, **47**, 331.

LEMAÎTRE, G., 1931: *Nature*, **128**, 704.

—— and VALLARTA, M. S., 1933: 'On Compton's Latitude Effect of Cosmic Radiation', *Phys. Rev.*, **43**, 87.

MALMFORS, K. G., 1945: 'Determination of Orbits in the Field of a Magnetic Dipole with Application to the Theory of the Diurnal Variation of Cosmic Radiation', *Ark. f. mat., astr. o. fysik*, **32 A**, No. 8.

MILLIKAN, R. A., NEHER, H. V., and PICKERING, W. H., 1942: 'A Hypothesis as to the Origin of Cosmic Rays and its Experimental Testing in India and Elsewhere', *Phys. Rev.*, **61**, 397.

MILLIKAN, R. A., NEHER, H. V., and PICKERING, W. H., 1943: 'Further Tests of the Atom-annihilation Hypothesis as to the Origin of the Cosmic Rays', *Phys. Rev.*, **63**, 234.

SCHEIN, M., JESSE, W. P., and WOLLAN, E. O., 1941: 'The Nature of the Primary Cosmic Radiation and the Origin of the Mesotron', ibid. **59**, 615; 'The Production of Mesotrons by Ionising Cosmic Rays', ibid. **59**, 930.

SWANN, W., 1933: 'A Mechanism of Acquirement of Cosmic-ray Energies by Electrons', *Phys. Rev.*, **43**, 217.

VALLARTA, M. S., 1937: 'Cosmic Rays and the Magnetic Moment of the Sun', *Nature*, **139**, 839.

NOTE

The year which has elapsed since this was written has changed the outlook on C.R. in many respects. High-altitude measurements by the photographic method, especially by Bradt and Peters (*Phys. Rev.* 1948), have shown that the primary radiation consists of atomic nuclei mixed in about the same proportions as in interstellar matter. This gives further support to the view that C.R. has been accelerated by electric fields in interstellar space.

Teller and Richtmeyer (*Phys. Rev.* 1949) have proposed that there are magnetic fields of the order 10^{-5} gauss in interstellar space, so that the radius of curvature of C.R. paths is small compared with interstellar distances. Hence C.R. should be a *local solar phenomenon*. This alternative to the universal and galactic hypotheses discussed in 7.4 makes the intensity problem much easier. The interstellar fields may be generated by magneto-hydrodynamic waves. A mechanism connecting the generation of C.R. with the magnetic storm effect (see 7.3) has been proposed (Alfvén, *Phys. Rev.* 1949). On the other hand, Fermi (*Phys. Rev.* 1949) has put forward a theory according to which C.R. is accelerated in interstellar space by the variable magnetic fields of magneto-hydrodynamic waves.

CHAPTER V. The formula for the Joule damping of a magneto-hydrodynamic wave was derived under the assumption of an isotropic conductivity and is not applicable when $\sigma_{\perp} \neq \sigma_{\parallel}$. In this case a circularly polarized wave can pass with a damping corresponding to $\sigma_{\parallel}$. This should have been considered in §§ 5.23, 5.43, and 5.51. As the circular polarization introduces an additional gravitational damping, the final result may be the same, at least qualitatively.

INDEX OF AUTHORS

INDEX OF SUBJECTS

PRINTED IN
GREAT BRITAIN
AT THE
UNIVERSITY PRESS
OXFORD
BY
CHARLES BATEY
PRINTER
TO THE
UNIVERSITY

COSMICAL ELECTRO-DYNAMICS

ALFVÉN

OXFORD